

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (uspro)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 04:26:11 ; Search time 59 seconds
(without alignments)
3701.657 Million cell updates/sec

Title: US-09-269-874A-3

Perfect score: 8424

Sequence: 1 MKIIFFLCSFLFIINTQCV.....SNFLGISFLILMLLYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database :

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8418	99.9	1639	19 AAW54145	P. falciparum synt
2	8139.5	96.6	1654	6 AAP50777	Sequence of the P1
3	1881	22.3	355	20 AAY09372	Merozoite surface
4	1881	22.3	355	20 AAY05832	Merozoite surface
5	1881	22.3	361	20 AAY09373	Merozoite surface
6	1881	22.3	361	20 AAY05833	Merozoite surface
7	1869	22.2	376	20 AAY09374	Modified merozoite
8	1869	22.2	376	20 AAY05834	Modified merozoite
9	1333	15.8	262	22 AAB37610	Merozoite surface
10	1165.5	13.8	375	22 AAB83926	A major merozoite

Polypeptide p190-3	355	10.8	913	9 AAP80547	Polypeptide p190-3
31-1 Repeated Dele	219	10.8	910.5	9 AAP82523	31-1 Repeated Dele
Polypeptide p190-1	219	10.7	905	9 AAP80544	Polypeptide p190-1
Storage-specific	244	9.4	794	6 AAP50304	Storage-specific
Polypeptide p190-2	151	8.0	677	9 AAP80546	Polypeptide p190-2
PfMSP1(p19)A prote	116	7.6	637	18 AAW36103	PfMSP1(p19)A prote
PfMSP1(p19)A prote	116	7.6	637	18 AAW22592	PfMSP1(p19)A prote
Plasmodium falcipa	115	6.9	584	23 AAM50527	Plasmodium falcipa
Merozoite surface	108	6.6	560	22 AAB37609	Merozoite surface
Plasmodium falcipa	1979	6.6	559.5	21 AAB18171	Plasmodium falcipa
Merozoite surface	96	6.6	556	22 AAB37608	Merozoite surface
PfMSP1(p19)S prote	127	6.4	539	18 AAW22593	PfMSP1(p19)S prote
PfMSP1(p19)S prote	127	6.4	539	18 AAW36102	PfMSP1(p19)S prote
Merozoite apical-en	1254	5.4	455.5	11 AAR07503	Merozoite apical-en
Merozoite apical-e	1254	5.4	455.5	18 AAW24575	Merozoite apical-e
Plasmodium falcipa	2485	5.4	453.5	21 AAB18172	Plasmodium falcipa
S. epidermidis ope	5024	5.3	450	22 AAG82935	S. epidermidis ope
Staphylococcus epi	10182	5.3	448	23 ABF38314	Staphylococcus epi
Human human diagno	2633	5.3	445	22 ABG06505	Human human diagno
Plasmodium polypeptide	2688	5.3	444.5	22 AAM40883	Plasmodium polypeptide
Human polypeptide	3973	5.3	442.5	21 AAB18253	Human polypeptide
Plasmodium falcipa	2663	5.2	441	22 AAM39097	Plasmodium falcipa
Novel human diagno	980	5.1	431	21 AAB18294	Novel human diagno
Plasmodium vivax m	2017	5.1	425.5	22 AAG06301	Plasmodium vivax m
Storage-specific	206	5.0	424.5	22 AAU00669	Storage-specific
Plasmodium falcipa	102	5.0	424	16 AAP50303	Plasmodium falcipa
Plasmodium falcipa	2010	4.9	414	21 AAB18218	Plasmodium falcipa
Plasmodium falcipa	1516	4.9	411.5	21 AAB18195	Plasmodium falcipa
P. falciparum telo	2539	4.8	408.5	21 AAB18198	P. falciparum telo
Plasmodium falcipa	2184	4.8	402.5	22 AAE00425	Plasmodium falcipa
Amino acid sequenc	2013	4.7	399	20 AAB18265	Amino acid sequenc
Plasmodium falcipa	2954	4.7	393	21 AAU01632	Plasmodium falcipa
Plasmodium falcipa	2295	4.6	388.5	20 AAB18180	Plasmodium falcipa
Staphylococcus epi	4134	4.6	387.5	21 AAB31946	Staphylococcus epi
Plasmodium falcipa	3696	4.6	387	23 ABP40235	Plasmodium falcipa
Plasmodium falcipa	1308	4.6	383.5	21 AAB18167	Plasmodium falcipa
Plasmodium falcipa	2573	4.6	383.5	21 AAB18234	Plasmodium falcipa
Human polypeptide	2206	4.5	379.5	21 AAB18254	Human polypeptide
Plasmodium falcipa	1780	4.5	375	22 AAB38681	Plasmodium falcipa
Nucleolar/endosoma	2500	4.4	372	21 AAB18272	Nucleolar/endosoma
Recombinant protei	1411	4.4	370	17 AAU02258	Recombinant protei
B. burgdorferi ant	1807	4.4	369.5	22 AAB85697	B. burgdorferi ant
Human polypeptide	1087	4.4	369	20 AAY19935	Human polypeptide
Plasmodium falcipa	1119	4.4	369	20 AAY19934	Plasmodium falcipa
Drosophila melanog	2380	4.3	368.5	22 AAM40467	Drosophila melanog
Plasmodium falcipa	2346	4.3	365.5	22 AAB63519	Plasmodium falcipa
Plasmodium falcipa	1712	4.3	363	21 AAB18205	Plasmodium falcipa
Novel human diagno	1714	4.3	360.5	21 AAB18275	Novel human diagno
Plasmodium falcipa	2415	4.3	359	22 ABG20279	Plasmodium falcipa
Novel human diagno	1817	4.2	357.5	21 AAB18255	Novel human diagno
Kinetochore protei	2400	4.2	356.5	22 ABG20278	Kinetochore protei
Drosophila melanog	3248	4.2	356.5	17 AAR99795	Drosophila melanog
Putative P. abyss	1690	4.2	354	22 ABB61144	Putative P. abyss
Human mitosisin amin	2482	4.2	354	16 AAR72826	Human mitosisin amin
Plasmodium falcipa	2482	4.2	354	19 AAW23996	Plasmodium falcipa
Plasmodium falcipa	1802	4.2	350.5	21 AAB18217	Plasmodium falcipa
Human alpha-myosin	1817	4.1	345	21 AAB18301	Human alpha-myosin
Drosophila melanog	1939	4.1	343	23 ABB77096	Drosophila melanog
Recombinant protei	1489	4.1	342.5	22 ABB59948	Recombinant protei
Human novel cytoki	2028	4.0	341	22 AAB85698	Human novel cytoki
Rattus norvegicus	2816	4.0	340	22 AAU68572	Rattus norvegicus
H. pylori GHPO 175	1886	4.0	338	19 AAW54241	H. pylori GHPO 175
Drosophila melanog	1743	4.0	336.5	22 ABB62322	Drosophila melanog
Sequence encoded b	2067	4.0	336.5	22 ABB71125	Sequence encoded b
	3685	4.0	335	10 AAP90373	

84	334.5	4.0	2918	22	ABG27218	Novel human diagno	157	298	3.5	3135	21	AAB18223	Plasmodium falcipa
85	334	4.0	1392	20	AAV06999	Restin protein seq	158	296.5	3.5	1149	23	ABB77620	AmEPV P4a core pro
86	333	4.0	1931	22	ABB61012	Drosophila melanog	159	296	3.5	1663	15	AAR46608	Plasmodium falcipa
87	332	3.9	976	22	AAG66581	Human SCP-1 mutain	160	295	3.5	1588	15	AAR46605	Malarial PfEMP3 ep
88	330	3.9	3685	10	AAP90290	Human Duchenne mus	161	295	3.5	2637	22	AAU31618	Novel human secret
89	330	3.9	7201	22	ABB71136	Drosophila melanog	162	295	3.5	4097	22	ABB57815	Drosophila melanog
90	329.5	3.9	1558	21	AAB18324	Plasmodium falcipa	163	294.5	3.5	1193	21	AAB18306	Plasmodium falcipa
91	329.5	3.9	2056	22	ABB59344	Drosophila melanog	164	294.5	3.5	1844	21	AAB18250	Plasmodium falcipa
92	329	3.9	996	22	AAU33755	Staphylococcus aur	165	294	3.5	1388	23	AAU79590	Human kinesin moto
93	329	3.9	1009	22	AAU36548	Staphylococcus aur	166	294	3.5	1388	23	AAE14400	Human kinesin supe
94	329	3.9	1346	21	AAB18236	Plasmodium falcipa	167	294	3.5	1786	14	AAR41043	CD4-EBAI75 fuson
95	329	3.9	2441	21	AAB18161	Plasmodium falcipa	168	293.5	3.5	1010	23	ABP40333	Staphylococcus epi
96	328	3.9	2496	21	AAB18222	Plasmodium falcipa	169	293.5	3.5	1084	23	ABB97247	Novel human protei
97	327.5	3.9	2172	22	ABB67958	Drosophila melanog	170	293	3.5	2391	15	AAR55694	Carbamoyl-phosphat
98	324.5	3.9	2096	22	AAB41592	Human ORFX ORF1356	171	293	3.5	3444	22	ABG17495	Novel human diagno
99	323.5	3.8	2139	22	AAB47278	Novel human diagno	172	292.5	3.5	1123	21	AAI58277	Heliothis armigera
100	323.5	3.8	2143	22	ABG01716	Human 160kD mediat	173	292.5	3.5	2025	22	AAU34207	Staphylococcus aur
101	322.5	3.8	1427	12	AAR10534	S. epidermidis ope	174	292.5	3.5	2986	22	AAU31620	Novel human secret
102	320	3.8	1095	22	AGB83030	Human polypeptide	175	292	3.5	1341	22	AAB84702	Amino acid sequenc
103	320	3.8	1988	22	RAM40999	Human polypeptide	176	292	3.5	5795	22	AAU37017	Staphylococcus aur
104	320	3.8	1988	22	RAM41000	Human polypeptide	177	291	3.5	1014	23	ABG61923	Prostate cancer-as
105	320	3.8	2779	22	ABG62371	Drosophila melanog	178	291	3.5	1327	23	ABBS7163	Mouse ischaemic co
106	319.5	3.8	1279	22	AGB3047	S. epidermidis ope	179	291	3.5	1960	22	AAU78854	Human protein SEQ
107	319.5	3.8	2659	23	AAU75489	S. aureus antigeni	180	291	3.5	2437	22	AAU34338	Staphylococcus aur
108	319	3.8	1436	21	AAB18199	Plasmodium falcipa							
109	318	3.8	1090	23	ABB54799	Lactococcus lactis							
110	317.5	3.8	1121	21	AAB18241	Plasmodium falcipa							
111	316.5	3.8	3158	22	AAU37018	Staphylococcus aur							
112	316	3.8	6815	22	ABB66811	Drosophila melanog							
113	315.5	3.7	1948	22	ABG21233	Novel human diagno							
114	314.5	3.7	1151	23	ABP38397	Staphylococcus epi							
115	312.5	3.7	1833	22	ABB71141	Drosophila melanog							
116	312.5	3.7	2048	22	AAU40027	Human polypeptide							
117	312	3.7	1409	22	ABB65706	Novel signal pathw							
118	312	3.7	2835	23	AAU75100	Staphylococcus epi							
119	310	3.7	1211	23	ABP39975	Novel human diagno							
120	310	3.7	1526	22	ABG07893	Streptococcus poly							
121	308.5	3.7	2059	23	ABP25711	Plasmodium falcipa							
122	308	3.7	696	21	AAB18181	Plasmodium falcipa							
123	308	3.7	931	22	AAU79504	Human protein SEQ							
124	308	3.7	1145	22	AGB2169	S. epidermidis ope							
125	308	3.7	1312	18	AAU22775	Human RAD50. Homo							
126	308	3.7	6281	22	AAU22775	Staphylococcus aur							
127	307.5	3.7	795	23	ABB77430	Human tumour marke							
128	307	3.6	54	14	AAR41356	MSPIEGF2A EGF2-lik							
129	307	3.6	1312	19	AAU71295	Human homologue of							
130	306.5	3.6	1211	22	ABB64663	Drosophila melanog							
131	306	3.6	1247	21	AAB18215	Plasmodium falcipa							
132	304.5	3.6	1182	21	AAB18288	Plasmodium falcipa							
133	304	3.6	990	22	AAU78520	Human protein SEQ							
134	304	3.6	1398	21	AAB18292	Plasmodium falcipa							
135	303.5	3.6	1038	22	AGB67416	Amino acid sequenc							
136	303.5	3.6	1975	22	ABB62094	Drosophila melanog							
137	303.5	3.6	2434	22	AAU34339	Staphylococcus aur							
138	303	3.6	1184	22	AAU33733	Staphylococcus aur							
139	303	3.6	1188	22	AAU36532	Staphylococcus aur							
140	303	3.6	1192	21	AAB18165	Plasmodium falcipa							
141	301	3.6	1046	23	ABB54658	Lactococcus lactis							
142	301	3.6	1355	22	ABG00331	Novel human diagno							
143	301	3.6	1626	22	ABG20974	Novel human diagno							
144	299	3.5	1166	21	AAB18268	Plasmodium falcipa							
145	299	3.5	1521	21	AAG39235	Arabidopsis thalia							
146	299	3.5	1528	21	AAG39234	Arabidopsis thalia							
147	299	3.5	1562	21	AAG39233	Arabidopsis thalia							
148	299	3.5	1703	21	AAG36714	Arabidopsis thalia							
149	299	3.5	1710	21	AAG36713	Arabidopsis thalia							
150	299	3.5	1744	21	AAG36712	Arabidopsis thalia							
151	298.5	3.5	2710	17	AAR5016	C. difficile toxin							
152	298.5	3.5	3329	19	AAW68387	Clostridium diffic							
153	298.5	3.5	3329	21	AAU56029	Human BRCA2 tumour							
154	298	3.5	1188	21	AAB18183	Plasmodium falcipa							
155	298	3.5	2194	22	AAU40114	Human polypeptide							
156	298	3.5	3135	15	AAR57474	P. falciparum tran							

ALIGNMENTS

RESULT 1

AAW54145
 ID AAW54145 standard; Protein; 1639 AA.
 XX AC AAW54145;
 XX DT 23-SEP-1998 (first entry)
 XX DE 23-SEP-1998 (first entry)
 XX KW P. falciparum synthetic gpl90 protein.
 XX KW gpl90; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 XX OS monoclonal antibody; passive immunisation; parasite.
 XX OS Plasmodium falciparum.
 XX OS Synthetic.
 XX PN WO9814583-A2.
 XX PD 09-APR-1998.
 XX PF 02-OCT-1997; 97WO-EP05441.
 XX PR 02-OCT-1996; 96DE-4040817.
 XX PA (BUJA/) BUJARD H.
 XX FI Bujard H, Pan W, Tolle R;
 XX DR WPI: 1998-240088/21.
 XX DR N-PSDB; RAV21451, RAV35363.
 XX PT Recombinant production of complete gpl90/MSP-1 Plasmodium surface
 XX PT protein - useful in anti-malaria vaccines, also stabilising genes by
 XX PS reducing their AT content
 XX PS Example 1; Fig 3c; 48pp; German.
 XX CC This sequence represents a modified Plasmodium falciparum gpl90/MSP-1
 XX CC (merozoite surface) protein. The gene encoding this protein has been
 XX CC stabilised by reducing the AT content of the nucleotide sequence. Such a
 XX CC protein is useful in vaccines against malaria or for producing monoclonal
 XX CC antibodies (for passive immunisation). The complete gpl90 protein can now
 XX CC be produced outside the parasite and has, at least over extended regions,
 XX CC the native pattern of folding. Larger amounts of the protein can be

CC "produced recombinantly than would be possible using the parasites as
CC source.

XX
SQ Sequence 1639 AA;

Query Match 99.9%; Score 8418; DB 19; Length 1639;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKIIFFLCSEFFIINTQCETHESYQELVKLEALEDAVLGTGSLFOKEKMWLNCEGTSCT 60
Dy 1 MKIIFFLCSEFFIINTQCETHESYQELVKLEALEDAVLGTGSLFOKEKMWLNCEGTSCT 60
Qy 61 AVTTSTPGSGSVASGGSGSVASGGSVASGGSVASGGSGGNSRRTPNSDSSDS 120
Dy 61 AVTTSTPGSGSVASGGSGSVASGGSVASGGSVASGGSGGNSRRTPNSDSSDS 120
Qy 121 DAKSYADLKURVRYLLTIKELKYPQFLDTNHNMLTLCNIDHGFYKYLIDGYEENELLYK 180
Dy 121 DAKSYADLKURVRYLLTIKELKYPQFLDTNHNMLTLCNIDHGFYKYLIDGYEENELLYK 180
Qy 181 LNFYFDLLRAKLDNDVCANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNIDKONVGKMD 240
Dy 181 LNFYFDLLRAKLDNDVCANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNIDKONVGKMD 240
Qy 241 YIKNKKTTIENINELIEESKKTIDKNKNATKEEKKKLYQAQYDLSYNNKQLEEAHNLIS 300
Dy 241 YIKNKKTTIENINELIEESKKTIDKNKNATKEEKKKLYQAQYDLSYNNKQLEEAHNLIS 300
Qy 301 VLEKRIDLTKKNENIKELLOKINIEKNPPANSNGTNTLLDNKKTEEHEKEIKETAKT 360
Dy 301 VLEKRIDLTKKNENIKELLOKINIEKNPPANSNGTNTLLDNKKTEEHEKEIKETAKT 360
Qy 361 IKFNIDSLFTDPLEYLLREKKNKIDISAKVETKESTEPENYNGVTPYPLSYNDINNAL 420
Dy 361 IKFNIDSLFTDPLEYLLREKKNKIDISAKVETKESTEPENYNGVTPYPLSYNDINNAL 420
Qy 421 NELNSFGDLINPFDTYKEPSKNIYTDNERKFTINEIKEKIEKKKIESDKKSYEDRSKS 480
Dy 421 NELNSFGDLINPFDTYKEPSKNIYTDNERKFTINEIKEKIEKKKIESDKKSYEDRSKS 480
Qy 481 LNDITKEYKLLNIDYDSKFNNDIDLNFEKMGKRYSYKVEKLTHTHTFASYSKHN 540
Dy 481 LNDITKEYKLLNIDYDSKFNNDIDLNFEKMGKRYSYKVEKLTHTHTFASYSKHN 540
Qy 541 EKLTKALKYMEDYSLRNIVVEKELKYKKNLISKIENIEIETLVENIKKDEQOLFETTKTD 600
Dy 541 EKLTKALKYMEDYSLRNIVVEKELKYKKNLISKIENIEIETLVENIKKDEQOLFETTKTD 600
Qy 601 ENKPDERILEYSDIVKVOQVKVLLMNNKIDELKKTQLILKNVELKHNIHVPSNYKQENKQE 660
Dy 601 ENKPDERILEYSDIVKVOQVKVLLMNNKIDELKKTQLILKNVELKHNIHVPSNYKQENKQE 660
Qy 661 PYLIVLVLKKEIDKLVKMPKVESLINEEKNKIKTEGSDNSEPSTEIGEITQATTKPGQ 720
Dy 661 PYLIVLVLKKEIDKLVKMPKVESLINEEKNKIKTEGSDNSEPSTEIGEITQATTKPGQ 720
Qy 721 AGSALBGDSVQAQEQKQAPVPVPVPEAKAQVTPPAPVNNKNTENVSKLDYLEKLYE 780
Dy 721 AGSALBGDSVQAQEQKQAPVPVPVPEAKAQVTPPAPVNNKNTENVSKLDYLEKLYE 780
Qy 781 FLNTSYCHKYILVSHSTMNEKILQYKTKTEESKLSGCDPLDLFNQNNIPVMYSMF 840
Dy 781 FLNTSYCHKYILVSHSTMNEKILQYKTKTEESKLSGCDPLDLFNQNNIPVMYSMF 840
Qy 841 DSLNNSLSQLFMEIYKEMVCNLYKLDNDKIKNLBEAKKVSTSVKTLSSSSMQPLSLT 900
Dy 841 DSLNNSLSQLFMEIYKEMVCNLYKLDNDKIKNLBEAKKVSTSVKTLSSSSMQPLSLT 900
Qy 901 PQDKPEVSANDDTSHSTNLNNSLKLFENILSLGKNKIYQELIGQKSSNFYEKILKDS 960
Dy 901 PQDKPEVSANDDTSHSTNLNNSLKLFENILSLGKNKIYQELIGQKSSNFYEKILKDS 960

Qy 961 TFYNESFTNFVKSKADDSINSLNDESKRKKLEEDINKLKTQLQSLFSLYNNKYKLEKLERLFD 1020
Dy 961 TFYNESFTNFVKSKADDSINSLNDESKRKKLEEDINKLKTQLQSLFSLYNNKYKLEKLERLFD 1020
Qy 1021 KKTIVGKYKMQIKKLTLLKEQLESKLSNLPKHVLFQNFVSFFNKKKEAEIAETENTLEN 1080
Dy 1021 KKTIVGKYKMQIKKLTLLKEQLESKLSNLPKHVLFQNFVSFFNKKKEAEIAETENTLEN 1080
Qy 1081 TKILLKHVGLVYNGESSPLKTLSEESTQEDNYASLENFKVLSKLEGLKDNLNLEK 1140
Dy 1081 TKILLKHVGLVYNGESSPLKTLSEESTQEDNYASLENFKVLSKLEGLKDNLNLEK 1140
Qy 1141 KKLSSLSGLHLLIAELKEVINKNKNYTCNSPSENNTDVNNALESYKFKLPEGTDVATVVS 1200
Dy 1141 KKLSSLSGLHLLIAELKEVINKNKNYTCNSPSENNTDVNNALESYKFKLPEGTDVATVVS 1200
Qy 1201 ESGSDTLQEQSQPKKPASTHVGAESNTITTSQNVDDVDDVIIIVPFGSEEDYDDLGOVV 1260
Dy 1201 ESGSDTLQEQSQPKKPASTHVGAESNTITTSQNVDDVDDVIIIVPFGSEEDYDDLGOVV 1260
Qy 1261 TGEAVTPSVTDNLSKLTENEVEVLYLKLPLAGVYKSLKQLENNVMTFNVVKDILNSRFN 1320
Dy 1261 TGEAVTPSVTDNLSKLTENEVEVLYLKLPLAGVYKSLKQLENNVMTFNVVKDILNSRFN 1320
Qy 1321 KRENFKNVLESDDLIPYKDLTSSNVVYKDPYKFLNKEKRDKFLSSYNYIKOSIDPDIINFAN 1380
Dy 1321 KRENFKNVLESDDLIPYKDLTSSNVVYKDPYKFLNKEKRDKFLSSYNYIKOSIDPDIINFAN 1380
Qy 1381 DVLGYKYLSEKYSKDLSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVHLE 1440
Dy 1381 DVLGYKYLSEKYSKDLSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVHLE 1440
Qy 1441 AKVLNYYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDYNNHNLTKFLS 1500
Dy 1441 AKVLNYYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDYNNHNLTKFLS 1500
Qy 1501 TGMVFENLAKTVLSNLLDGNLQGLMNTSQHCYKQCPQNSGCGPRHLDERECKCLLNK 1560
Dy 1501 TGMVFENLAKTVLSNLLDGNLQGLMNTSQHCYKQCPQNSGCGPRHLDERECKCLLNK 1560
Qy 1561 QEGDKCVENPNTCENNGGCDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFCSSS 1620
Dy 1561 QEGDKCVENPNTCENNGGCDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFCSSS 1620
Qy 1621 NFLGISFLLIIMLLIYSFI 1639
Dy 1621 NFLGISFLLIIMLLIYSFI 1639
RESULT 2
AAP50777
ID AAP50777 standard; Protein; 1654 AA.
XX AAP50777;
AC AAP50777;
DT 30-SEP-1991 (first entry)
XX
DE Sequence of the P195 protein of Plasmodium falciparum.
XX Malaria vaccine; epitope; antigen; immunogen.
XX Plasmodium falciparum.
XX EP154454-A.
XX 11-SEP-1985.
XX 21-FEB-1985; 85EP-0301173.
XX 26-SEP-1984; 84GB-0024340.
XX 22-FEB-1984; 84GB-0004692.
XX 21-FEB-1985; 85GB-0004429.
XX


```
XX  Merozoite surface protein MSP-1-42.
DE  MSP-1: merozoite surface protein; malaria; vaccine;
KW  protein engineering; protein expression; codon usage;
KW  transgenic animal.
XX  Plasmodium falciparum.
OS  W09920774-A2.
XX  W09920774-A2.
XX  29-APR-1999.
XX  20-OCT-1998; 98WO-US22226.
XX  15-MAY-1998; 98US-0085649.
XX  20-OCT-1997; 97US-0062592.
XX  (GENZ ) GENZYME TRANSGENICS CORP.
FA  Chen LH, Meade H;
XX  WPI: 1999-288313/24.
XX  P-PSDB: AAX56008.
XX  Modified malarial protein for use in anti-malarial vaccines
XX  Example; Fig 1; 35pp; English.
XX  The present sequence represents a 42 kDa C-terminal portion of
CC  malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC  important target for the development of a vaccine against
CC  Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
CC  has been modified (see AAX56008) compared to the native sequence (see
CC  AAX56009) such that 306 nucleotide positions have been replaced to
CC  lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC  instability motifs while maintaining the same protein amino acid
CC  sequence. These alterations allow MSP-1-42 to be expressed in
CC  mammalian cell culture and in transgenic mice. Native MSP-1-12
CC  is known to be difficult to express in cell culture systems.
CC  mammalian cell culture systems or in transgenic animals. The
CC  invention allows expression of MSP-1 protein in the milk of
CC  transgenic animals, and also provides a DNA vaccine comprising a
CC  vector containing the altered MSP-1-42 sequence.
XX  Sequence 355 AA;
SQ  Query Match 22.3%; Score 1881; DB 20; Length 355;
    Best Local Similarity 100.0%; Pred. No. 4.4e-75;
    Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1264 AVTPSVIDNILSKTENYEVLVYLPAGVYRSLKKOLENNVMTFNVNVKDIILNSRFNKRE 1323
Db 1 AVTPSVIDNILSKTENYEVLVYLPAGVYRSLKKOLENNVMTFNVNVKDIILNSRFNKRE 60
QY 1324 NFKNVLESDLIPYKDLTSSNVVVDYPYKFLNKRDKFLSSYNIKDSIDTDINFANDVL 1383
Db 61 NFKNVLESDLIPYKDLTSSNVVVDYPYKFLNKRDKFLSSYNIKDSIDTDINFANDVL 120
QY 1384 GYKILSEKYSDDLISIKKYINDKOGENEKYLPLFNNTIETLYKTVNDKIDLFVHLEAKV 1443
Db 121 GYKILSEKYSDDLISIKKYINDKOGENEKYLPLFNNTIETLYKTVNDKIDLFVHLEAKV 180
QY 1444 LNYTEKSNVVKIKELNYLTKTDKADFKKNNFVGIADLSTDYNNHLLTKFLSGM 1503
Db 181 LNYTEKSNVVKIKELNYLTKTDKADFKKNNFVGIADLSTDYNNHLLTKFLSGM 240
QY 1504 VFENLAKTVLSNLDGNLQGMINTSQHCVKQCPQNSGCFRHLDERECKLLNYKQEG 1563
Db 241 VFENLAKTVLSNLDGNLQGMINTSQHCVKQCPQNSGCFRHLDERECKLLNYKQEG 300
QY 1564 DKCVENPNPTCENNNGGCDADAKTEEDSGSNGKKITCECTKPDSPYPLFDGIFCS 1618
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 301 DKCVENPNPTCENNNGGCDADAKTEEDSGSNGKKITCECTKPDSPYPLFDGIFCS 355
RESULT 4
AAY05832
ID AAY05832 standard; Protein; 355 AA.
XX AC AAY05832;
XX XX 02-AUG-1999 (first entry)
XX DE Merozoite surface protein MSP-1-42.
XX KW MSP-1: merozoite surface protein; malaria; vaccine;
XX KW protein engineering; protein expression; codon usage;
XX KW transgenic animal.
XX OS Plasmodium falciparum.
XX PN W09920766-A2.
XX XX 29-APR-1999.
XX PF 20-OCT-1998; 98WO-US22225.
XX PR 15-MAY-1998; 98US-0085649.
XX PR 20-OCT-1997; 97US-0062592.
XX XX (GENZ ) GENZYME TRANSGENICS CORP.
XX PA Chen LH, Meade H;
XX PI WPI: 1999-302742/25.
XX DR N-PSDB: AAX25586.
XX XX New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX Disclosure; Fig 1; 43pp; English.
XX The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
CC has been modified (see AAX25586) compared to the native sequence (see
CC AAX25587) such that 306 nucleotide positions have been replaced to
CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC instability motifs while maintaining the same protein amino acid
CC sequence. These alterations allow MSP-1-42 to be expressed in
CC mammalian cell culture and in transgenic mice. The invention
CC provides modified recombinant nucleic acid sequences and methods for
CC increasing the mRNA levels and protein expression of proteins that
CC are difficult to express in cell culture systems, mammalian cell
CC culture systems or in transgenic animals. The preferred difficult
CC protein candidates for expression are those derived from lower
CC organisms such as parasites, bacteria and viruses that have DNA
CC coding sequences of high AT content or which have mRNA instability
CC motifs or rare codons relative to the recombinant expression system
CC to be used. The invention allows expression of MSP-1 protein in
CC the milk of transgenic animals, and also provides a DNA vaccine
CC comprising a vector containing the altered MSP-1-42 sequence.
XX SQ Sequence 355 AA;
    Query Match 22.3%; Score 1881; DB 20; Length 355;
    Best Local Similarity 100.0%; Pred. No. 4.4e-75;
    Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1264 AVTPSVIDNILSKTENYEVLVYLPAGVYRSLKKOLENNVMTFNVNVKDIILNSRFNKRE 1323
Db 1 AVTPSVIDNILSKTENYEVLVYLPAGVYRSLKKOLENNVMTFNVNVKDIILNSRFNKRE 60
QY 1324 NFKNVLESDLIPYKDLTSSNVVVDYPYKFLNKRDKFLSSYNIKDSIDTDINFANDVL 1383
Db 1324 NFKNVLESDLIPYKDLTSSNVVVDYPYKFLNKRDKFLSSYNIKDSIDTDINFANDVL 1383
```

Db 61 NFKNVLESLLPYKDLTSSNYVVKPYKFLNKEKRDKFLSSNYIKDSIDTDINFANDVL 120
QY 1384 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFLNNIETKYTVNDKIDLFVHLEAKV 1443
Db 121 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFLNNIETKYTVNDKIDLFVHLEAKV 180
QY 1444 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNNFVGTADLSTDYNNHNNLLTKFLSTGM 1503
Db 181 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNNFVGTADLSTDYNNHNNLLTKFLSTGM 240
QY 1504 VFNENLAKTVLSNLLDGNLQGMNLNISOHQCVKKQCPQNSGCFRHLDERECKCLLNYKQEG 1563
Db 241 VFNENLAKTVLSNLLDGNLQGMNLNISOHQCVKKQCPQNSGCFRHLDERECKCLLNYKQEG 300
QY 1564 DKCVENPNTCENNGGCCDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCS 1618
Db 301 DKCVENPNTCENNGGCCDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCS 355

RESULT 5
AAV09373
ID AAY09373 standard; Protein; 361 AA.
XX
AC AAY09373;
XX

DT 31-AUG-1999 (first entry)
XX

DE Merozoite surface protein MSP-1-42.
XX

KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX

OS Plasmodium falciparum.
XX

PN WO9920774-A2.
XX

XX 29-APR-1999.
PD

PF 20-OCT-1998; 98WO-US22226.
XX

XX 15-MAY-1998; 98US-0085649.
PR

PR 20-OCT-1997; 97US-0062592.
XX

XX (GENZ) GENZYME TRANSGENICS CORP.
PA

XX Chen LH, Meade H;
PI

XX WPI; 1999-288313/24.
DR

DR P-PSDB; AAX56009.
XX

XX Modified malarial protein for use in anti-malarial vaccines
PT

XX Example; Fig 2; 35pp; English.
PS

XX This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an

CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is

CC modified to include a 6xHis tag. A nucleic acid (see AAX56008)
CC encoding MSP-1-42 has been modified according to a method

CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals by reducing the AT content and removing

CC mRNA instability motifs. The invention allows expression of
CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA

CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC acid.
XX

XX Sequenced 361 AA;
SQ

Query Match 22.3%; Score 1881; DB 20; Length 361;
Best Local Similarity 100.0%; Pred. No. 4.5e-75;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1264 AVTPSVIDNLSKINENEYEVLYKPLAGVYRSLKKOLENNVMTFNVVKDILNSRFNKE 1323
Db 1 AVTPSVIDNLSKINENEYEVLYKPLAGVYRSLKKOLENNVMTFNVVKDILNSRFNKE 60
QY 1324 NFKNVLESLLPYKDLTSSNYVVKPYKFLNKEKRDKFLSSNYIKDSIDTDINFANDVL 1383
Db 61 NFKNVLESLLPYKDLTSSNYVVKPYKFLNKEKRDKFLSSNYIKDSIDTDINFANDVL 120
QY 1384 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFLNNIETKYTVNDKIDLFVHLEAKV 1443
Db 121 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFLNNIETKYTVNDKIDLFVHLEAKV 180
QY 1444 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNNFVGTADLSTDYNNHNNLLTKFLSTGM 1503
Db 181 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNNFVGTADLSTDYNNHNNLLTKFLSTGM 240
QY 1504 VFNENLAKTVLSNLLDGNLQGMNLNISOHQCVKKQCPQNSGCFRHLDERECKCLLNYKQEG 1563
Db 241 VFNENLAKTVLSNLLDGNLQGMNLNISOHQCVKKQCPQNSGCFRHLDERECKCLLNYKQEG 300
QY 1564 DKCVENPNTCENNGGCCDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCS 1618
Db 301 DKCVENPNTCENNGGCCDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCS 355

RESULT 6

AAV05833

ID AAY05833 standard; Protein; 361 AA.
XX

AC AAY05833;
XX

DT 02-AUG-1999 (first entry)
XX

DE Merozoite surface protein MSP-1-42.
XX

KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX

OS Plasmodium falciparum.
XX

XX WO9920766-A2.
PN

XX 29-APR-1999.
PD

XX 20-OCT-1998; 98WO-US22225.
PF

XX 15-MAY-1998; 98US-0085649.
PR

PR 20-OCT-1997; 97US-0062592.
XX

XX (GENZ) GENZYME TRANSGENICS CORP.
PA

XX Chen LH, Meade H;
PI

XX WPI; 1999-302742/25.
DR

DR N-PSDB; AAX25587.
XX

XX New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine

XX Disclosure; Fig 2; 43pp; English.
PS

XX This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an

CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is

CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
CC AAX25593) encoding MSP-1-42 have been modified according to a method

CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals. The invention provides modified

CC recombinant nucleic acid sequences and methods for increasing the

CC 'mRNA levels and protein expression of proteins that are difficult
 CC to express in cell culture systems, especially mammalian cell
 CC culture systems or in transgenic animals. The preferred difficult
 CC protein candidates for expression are those derived from lower
 CC organisms such as parasites, bacteria and viruses that have DNA
 CC coding sequences of high A/T content or which have mRNA instability
 CC motifs or rare codons relative to the recombinant expression system
 CC to be used. The invention allows expression of MSP-1 in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 nucleic acid.

XX SQ Sequence 361 AA;

Query Match 22.3%; Score 1881; DB 20; Length 361;
 Best Local Similarity 100.0%; Pred. No. 4.5e-75;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1264 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLNNVMTFNVNVKDIILNSRFNKR 1323

Db 1 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLNNVMTFNVNVKDIILNSRFNKR 60

Qy 1324 NFKNVLESDLIPYKDLTSSNVVVDPKYFLNKRDRKFLSSYNYIKDSIDTDINFANDVL 1383

Db 61 NFKNVLESDLIPYKDLTSSNVVVDPKYFLNKRDRKFLSSYNYIKDSIDTDINFANDVL 120

Qy 1384 GYYKILSEKYSKDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 1443

Db 121 GYYKILSEKYSKDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 180

Qy 1444 LNYTYEKSNEVKIKELNYLKTIDQKLADFKKNNFVGADLSTDYNNHNLTKFLSTGM 1503

Db 181 LNYTYEKSNEVKIKELNYLKTIDQKLADFKKNNFVGADLSTDYNNHNLTKFLSTGM 240

Qy 1504 VFENLAKTVLSNLDGNLQGLMNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNKQEG 1563

Db 241 VFENLAKTVLSNLDGNLQGLMNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNKQEG 300

Qy 1564 DKCVENPNTCNENGGCDADAKCTEEDSGNGKKITCECTKPDSPYPLFDGIFCS 1618

Db 301 DKCVENPNTCNENGGCDADAKCTEEDSGNGKKITCECTKPDSPYPLFDGIFCS 355

RESULT 7

AA09374

ID AA09374 standard; Protein; 376 AA.

XX AC AA09374;

XX 31-AUG-1999 (first entry)

XX Modified merozoite surface protein MSP-1-42.

XX MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.

XX Plasmodium falciparum.
 OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..15

FT Peptide /note= "beta-casein signal peptide"

FT Peptide 371..376

FT Peptide /note= "6xHis tag"

FT Misc-difference 197

FT Misc-difference /note= "Asn in native MSP-1-42 (N181Q mutation)"

FT Misc-difference 278

FT Misc-difference /note= "Asn in native MSP-1-42 (N262Q mutation)"

XX WO9920774-A2.

XX 29-APR-1999.

XX

PF 20-OCT-1998; 98WO-US22226.

XX 15-MAY-1998; 98US-0085649.

PR 20-OCT-1997; 97US-0062592.

XX (GENZ) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI; 1999-288313/24.

DR P-PSDB; AAX56008.

XX Modified malarial protein for use in anti-malarial vaccines

XX Example; Fig 11; 35pp; English.

XX The present sequence represents a 42 kDa C-terminal portion of

CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an

CC important target for the development of a vaccine against

CC Plasmodium falciparum. The sequence has been modified to include

CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis

CC tag. In addition, N181Q and N262Q mutations have been introduced to

CC eliminate 2 N-glycosylation sites. These modifications allow the

CC MSP-1-42 protein to be expressed in the milk of transgenic mice.

XX

SQ Sequence 376 AA;

Query Match 22.2%; Score 1869; DB 20; Length 376;

Best Local Similarity 99.4%; Pred. No. 1.6e-74;

Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1264 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLNNVMTFNVNVKDIILNSRFNKR 1323

Db 16 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLNNVMTFNVNVKDIILNSRFNKR 75

Qy 1324 NFKNVLESDLIPYKDLTSSNVVVDPKYFLNKRDRKFLSSYNYIKDSIDTDINFANDVL 1383

Db 76 NFKNVLESDLIPYKDLTSSNVVVDPKYFLNKRDRKFLSSYNYIKDSIDTDINFANDVL 135

Qy 1384 GYYKILSEKYSKDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 1443

Db 136 GYYKILSEKYSKDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 195

Qy 1444 LNYTYEKSNEVKIKELNYLKTIDQKLADFKKNNFVGADLSTDYNNHNLTKFLSTGM 1503

Db 196 LNYTYEKSNEVKIKELNYLKTIDQKLADFKKNNFVGADLSTDYNNHNLTKFLSTGM 255

Qy 1504 VFENLAKTVLSNLDGNLQGLMNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNKQEG 1563

Db 256 VFENLAKTVLSNLDGNLQGLMNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNKQEG 315

Qy 1564 DKCVENPNTCNENGGCDADAKCTEEDSGNGKKITCECTKPDSPYPLFDGIFCS 1618

Db 316 DKCVENPNTCNENGGCDADAKCTEEDSGNGKKITCECTKPDSPYPLFDGIFCS 370

RESULT 8

AA05834

ID AA05834 standard; Protein; 376 AA.

XX AC AA05834;

XX 02-AUG-1999 (first entry)

XX Modified merozoite surface protein MSP-1-42.

XX MSP-1; merozoite surface protein; malaria; vaccine;

KW protein engineering; protein expression; codon usage;

KW transgenic animal; mutant.

XX Plasmodium falciparum.

OS Synthetic.

XX


```
Qy 160 NIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPFNKIRANELDVLLK 219
      |||||
Db 149 NIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPFNKIRANELDVLLK 208
      |||||
Qy 220 LVFGYRKPLDNKIDNVGKMDYIKKNKKTNIENINELIEESKKTIDKNKNATKEEKKLY 279
      |||||
Db 209 LVFGYRKPLDNKIDNVGKMDYIKKNKKTNIENINELIEESKKTIDKNKNATKEEKKLY 268
      |||||
Qy 280 QAOYDLSTYINKOLEAHNLISVLEKRIIDLTKKNENIKELDKINEIKNPPPA 331
      |||||
Db 269 QAOYDLFTYINKOLEAHNLISVLEKRIIDLTKKNENIKELDKINEIKNPPPA 320
      |||||

RESULT 12
AAP82523
ID AAP82523 standard; protein; 219 AA.
XX
AC AAP82523;
XX
DT 12-DEC-1990 (first entry)
XX
DE 31-1 Repeated Delete.
XX
KW Malaria; stage-specific late schizont-merozoite antigen 31-1;
KW vaccine; protozoan parasite; repitope.
XX
OS Plasmodium falciparum.
XX
PN EP254862-A.
XX
PD 03-FEB-1988.
XX
PF 20-JUN-1987; 87EP-0108867.
XX
PR 26-JUN-1986; 86US-0879076.
XX
PA (BEHW ) BEHRINGWERKE AG.
XX
PI Shaw A, Humbert Y;
XX
DR WPI; 1988-030152/05.
DR N-PSDB; AAN82176.
XX
PT Peptide(s) for prodn. of antiprotozoal vaccines - comprising
PT surface antigen sequence with deleted repitope
XX
PS Disclosure; ; p; English.
XX
CC The sequence encodes a polypeptide which contains the unique
CC sequence of the surface antigen of P. falciparum, but lacks the
CC immunodominant repeat sequences which may allow the parasite to
CC evade and decoy the immune system. The DNA for the stage-
CC specific late schizont-merozoite antigen 31-1 in p31-1 (W08503725)
CC was modified to delete the entire sequence coding for the repitope.
CC The expression plasmid, 31-1 Repeated Delete, was introduced into
CC E.coli K12 to produce the modified peptide. The transformant has
CC been deposited in the Deutsche Sammlung von Mikroorganismen as
CC E.coli K12 (9SG936cl) (p31-1 Repeat Delete). The peptide was used
CC in compsns. to treat and prevent protozoan parasitic infections.
XX
SQ Sequence 219 AA;
      Query Match 10.8%; Score 910.5; DB 9; Length 219;
      Best Local Similarity 76.2%; Pred. No. 8.6e-33;
      Matches 182; Conservative 4; Mismatches 8; Indels 45; Gaps 1;

Qy 26 QELVKKLEALEDAVLTYSLFQEKWVLNEGTSVAVTTSTPGSKGVSAGSGGVSAG 85
      |||||
Db 6 QELVKKLEALEDAVLTYSLFQEKWVLNEGTSVAVTTSTPGSKGVSAGSGGVSAG 45
      |||||

Qy 86 GSVASGGVSAGSGGVSAGSGNSRTNPSDSSSDSAKSYADLKHVRVNYLLTIKELKYP 145
      |||||
```

```
Db 46 -----TNPSDSSSDSAKSYADLKHVRVNYLLTIKELKYP 80
Qy 146 OLFDLTNHLTLCDNIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPF 205
      :|||||
Db 81 ELFDLTNHLTLCDNIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPF 140
Qy 206 NLKIRANELDVLKLVFGYRKPLDNKIDNVGKMDYIKKNKKTNIENINELIEESKKTID 264
      |||||
Db 141 NLKILRNELDLVKLVFGYRKPLDNKIDNVGKMDYIKKNKKTNIENINELIEESKKTID 199

RESULT 13
AAP80544
ID AAP80544 standard; protein; 219 AA.
XX
AC AAP80544;
XX
DT 11-NOV-1990 (first entry)
XX
DE Polypeptide p190-1.
XX
KW Polypeptide p190-1; P.falciparum; merozoite; vaccine.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT Region 1..8
FT /label=affinity peptide
FT Region 184..219
XX
XX EP283829-A.
XX
PD 28-SEP-1988.
XX
PF 08-MAR-1988; 88EP-0103564.
XX
PR 19-MAR-1987; 87GB-0006599.
XX
PA (HOFF ) F. HOFFMANN-LA ROCHE & CO.
XX
PI Ulrich C, Gentz H, Takacs B;
XX
DR WPI; 1988-272339/39.
DR N-PSDB; AAN81148.
XX
PT New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
PT surface antigen precursor useful in vaccines, and encoding DNA sequences
XX
PS Claim 4; Page 26; 45pp; German.
XX
CC The polypeptide contains at least one epitope of the 190 kD precursor of
CC the major merozoite surface antigen of P.falciparum. The polypeptide is
CC immunogenic and can be used in vaccines to protect against malaria.
CC Presence of the affinity peptide allows the polypeptide to be purified
CC by affinity chromatography resins. Antibodies raised against the
CC polypeptide are useful for diagnosis and for purificn. by affinity
CC chromatography.
XX
SQ Sequence 219 AA;
      Query Match 10.7%; Score 905; DB 9; Length 219;
      Best Local Similarity 98.9%; Pred. No. 1.5e-32;
      Matches 174; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 156 TLCDNIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPFNKIRANELD 215
      |||||
Db 9 TLCDNIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPFNKIRANELD 68
Qy 216 VLKLVFGYRKPLDNKIDNVGKMDYIKKNKKTNIENINELIEESKKTIDKNKNATKEEK 275
      |||||
Db 69 VLKLVFGYRKPLDNKIDNVGKMDYIKKNKKTNIENINELIEESKKTIDKNKNATKEEK 128
Qy 276 KLYQAQYDLSTYINKOLEAHNLISVLEKRIIDLTKKNENIKELDKINEIKNPPPA 331
```

Db 129 KKLQAQYDLFIYNKQLEAHNLISVLEKRIIDTLKKNNIKELLDKINEIKNPPPA 184
RESULT 14
AAP50304
ID AAP50304 standard; Protein; 244 AA.
XX AC
XX AAP50304;
XX DT 22-OCT-1991 (first entry)
XX DE Storage-specific, late schizont merozoite malaria antigen.
XX KW Malaria; vaccine.
XX OS Plasmodium spp.
XX PN W08503725-A.
XX PD 29-AUG-1985.
XX PF 20-FEB-1985; 85WO-GB000072.
XX PR 20-FEB-1984; 84GB-0004378.
XX PA (BIOJ) BIOGEN NV.
XX PI Mach B, Perrin L, McGarvey M, Cheung A, Shaw A;
XX DR WPI; 1985-223371/36.
XX DR N-PSDB; AAN50355.
XX PT Prodn. of antigens of Plasmodium species, esp. of falciparum - by
PT recombinant DNA methods giving polypeptide(s) for protecting
PT against malaria or for diagnosis.
XX PS Disclosure; Fig 6; 49pp; English.
XX CC The sequence encodes a Plasmodium falciparum, Plasmodium vivax,
CC Plasmodium malariae and Plasmodium ovale antigen which may be used
CC in the diagnosis of malaria and as a vaccine against malaria.
XX SQ Sequence 244 AA;
Query Match 9.4%; Score 794; DB 6; Length 244;
Best Local Similarity 92.0%; Pred. No. 1.2e-27;
Matches 160; Conservative 3; Mismatches 5; Indels 6; Gaps 2;
Qy 1 MKIIFFLCSPFFIINTOCVTTHESYQELVKKLEALEDAVLTYGSLFQEKVMVNEGTS 60
Db 75 MKIIFFLCSPFFIINTOCVTTHESYQELVKKLEALEDAVLTYGSLFQEKVMVNEGTS 134
Qy 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSGNSRRTNPSDSDS 120
Db 135 AVTTSTPGSGSVTSGGGSGVA---SVASGG---SGSGSVASGGSGNSRRTNPSDSDS 188
Qy 121 DAKSYADLKHRVRLVLTIKELYPOLFDLTNHLMTLCDNIHGFYLDGYEEI 174
Db 189 DAKSYADLKHRVQVLYFTIKELYPOLFDLTNHLMTLCDNIHGFYLDGYEEI 242
RESULT 15
AAP80546
ID AAP80546 standard; protein; 151 AA.
XX AC
XX AAP80546;
XX DT 11-NOV-1990 (first entry)
XX DE Polypeptide p190-2b.
XX KW Polypeptide p190-2b; P.falciparum; merozoite; vaccine.

XX Plasmodium falciparum.
XX FH Key Location/Qualifiers
FT Region 1..5
FT /label-affinity peptide
FT Region 142..142
XX PN EP283829-A.
XX PD 28-SEP-1988.
XX PF 08-MAR-1988; 88EP-0103564.
XX PR 19-MAR-1987; 87GB-0006599.
XX PA (HOFF) F. HOFFMANN-LA ROCHE & CO.
XX PI Ulrich C, Gentz H, Takacs B;
XX DR WPI; 1988-272339/39.
XX DR N-PSDB; AAN81150.
XX PT New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
PT surface antigen precursor, useful in vaccines, and encoding DNA sequences
XX PS Claim 6; Page 27; 45pp; German.
XX CC The polypeptide contains at least one epitope of the 190 kD precursor of
CC the major merozoite surface antigen of P.falciparum. The polypeptide is
CC immunogenic and can be used in vaccines to protect against malaria.
CC Presence of the affinity peptide allows the polypeptide to be purified
CC by affinity chromatography resins. Antibodies raised against the
CC polypeptide are useful for diagnosis and for purificn. by affinity
CC chromatography.
XX SQ Sequence 151 AA;
Query Match 8.0%; Score 677; DB 9; Length 151;
Best Local Similarity 95.8%; Pred. No. 9.5e-23;
Matches 136; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
Qy 1069 AEIAETENTLENTKILLKHVYKNGESSPLKTLSEESIQTEDNYASLENFKVLSKL 1128
Db 6 AEIAETENTLENTKILLKHVYKNGESSPLKTLSEESIQTEDNYASLENFKVLSKL 65
Qy 1129 EGKLDNLNLEKKLSYLSGLHLLIAELKEVIKKNYTGNSPSNNTDVNNALSYKKF 1188
Db 66 EGKLDNLNLEKKLSYLSRGLHLLIAELKEVIKKNYTGNSPSNNTDVNNALSYKKF 125
Qy 1189 LPEGTDVATVVSSESGSDTLEQS 1210
Db 126 LPEGTDVATVVSSESGSDVLDQPS 147
RESULT 16
AAW36103
ID AAW36103 standard; Protein; 116 AA.
XX AC AAW36103;
XX DT 25-MAR-1998 (first entry)
XX DE PFMSP1(p19)A protein sequence.
XX KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Region 1..95

PA (GEOU) UNIV GEORGETOWN.

XX Davidson E, Nikodem D;
XX WPI; 2002-049444/06.
XX Novel vaccine for immunizing mammals against Plasmodium falciparum
PT infection, comprises a viral vector system expressing protein
PT corresponding to specific domain of major merozoite surface protein 1
PT of Plasmodium falciparum
XX
XX Disclosure; Page 4; 60pp; English.

XX The present sequence is that of immunogenic protein p115MSP-1, a
CC specific domain (amino acids 1002-1116) of the major merozoite
CC surface protein 1 (MSP-1) of the erythrocyte stage of the
CC West African Wellcome strain of Plasmodium falciparum. The
CC invention relates to a malaria vaccine comprising an expression
CC vector, preferably a defective Venezuelan equine encephalitis (VVEE)
CC viral vector system, which expresses p115MSP-1 or its immunogenic
CC fragment after administration to a patient. The p115MSP-1 protein
CC or fragment raises a humoral and/or cell-mediated response to the
CC erythrocytic merozoite malaria antigen, protecting the patient from
CC a subsequent malaria infection. The VVEE viral vector system
CC continues to express antigen in the patient for a period of days,
CC months or even years. The p115MSP-1 antigen may also include
CC a signal peptide and/or an anchor peptide sequence, which
CC enhances the immunogenic response of the patient.

XX Sequence 115 AA;

Query Match 6.9%; Score 584; DB 23; Length 115;

Best Local Similarity 100.0%; Pred. No. 8.1e-19;

Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 QLSFDLYNKYKLERLFDKKTGKYKMQIKKTLLEQLESKLSLNPKHVLQNFV 1061

Db 1 QLSFDLYNKYKLERLFDKKTGKYKMQIKKTLLEQLESKLSLNPKHVLQNFV 60

Qy 1062 FFKKKKEAIEATENTLTKLLKHYGLVYKNGESSPLKTLSEESIQTEDNY 1116

Db 61 FFKKKKEAIEATENTLTKLLKHYGLVYKNGESSPLKTLSEESIQTEDNY 115

RESULT 19

AAB37609

ID AAB37609 standard; Protein; 108 AA.

XX AAB37609;

DT 27-FEB-2001 (first entry)

DE Merozoite surface protein-119.

XX Merozoite surface protein; protazoacide; vaccine; malaria.

XX Plasmodium falciparum.

XX WO2000063245-A2.

XX 26-OCT-2000.

XX 20-APR-2000; 2000WO-GB01558.

XX 20-APR-1999; 99GB-0009072.

XX 13-MAY-1999; 99US-0311817.

XX 25-MAY-1999; 99CA-2271451.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiipibull C;

XX WPI; 2001-015762/02.

DR N-PSDB; AAC68977.

XX Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria
PT
XX Example 5; Fig 15; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal

CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is MSP-119 protein.

XX Sequence 108 AA;

Query Match 6.6%; Score 560; DB 22; Length 108;

Best Local Similarity 94.1%; Pred. No. 8.5e-18;

Matches 96; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1520 NLQGLMLTSQHCVKVKQPCQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNG 1579

Db 7 HIEGRHNIAHQCVKVKQPCQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNG 66

Qy 1580 GCDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFCSSN 1621

Db 67 GCDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFCSSN 108

RESULT 20

AAB18171

ID AAB18171 standard; Protein; 1979 AA.

XX AAB18171;

XX 07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:28.

XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;

XX antimalarial; malaria; protozoacide; infection; insecticide.

XX Plasmodium falciparum.

XX WO200025728-A2.

XX 11-MAY-2000.

XX 05-NOV-1999; 99WO-US26796.

XX 05-NOV-1998; 98US-0107131.

XX (HOFF/) HOFFMAN S.

XX (CARU/) CARUCCI D.

XX (GARD/) GARDNER M.

XX (VENT/) VENTER J C.

XX Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI; 2000-365347/31.

XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -

XX Disclosure; Page 70-75; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.

Also described are: (1) nucleotide sequences (II) encoding (I); and (2) vaccines against *P. falciparum* infection comprising (I) or (II). (I) and (II) are useful for the development of vaccines against *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with *P. falciparum*. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAAY70078 to AAA70287 and AAB18144 to AAB19352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

```

XX Merozoite surface protein-1.
DE Merozoite surface protein; protozoacide; vaccine; malaria.
KW Merozoite surface protein; protozoacide; vaccine; malaria.
XX Plasmodium falciparum.
OS Plasmodium falciparum.
XX WO200063245-A2.
XX 26-OCT-2000.
XX 20-APR-2000; 2000WO-GB01558.
XX 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX (MEDI-) MEDICAL RES COUNCIL.
XX Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX WPI; 2001-015762/02.
XX Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria.
XX Example 2; Page 48; 126pp; English.
XX The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The present sequence is the
CC wild-type MSP-1 protein. This sequence was used to generate the variants
CC of the present invention. The non-natural variants of the present
CC invention are useful for immunising a mammal against malaria, and can be
CC used to treat malaria.
XX Sequence 96 AA;
XX Query Match 6.6%; Score 556; DB 22; Length 96;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-17;
XX Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1526 NISQHCVKKQCPQNSGCFRHLDERECKLLNKKQSGDKCVENPNTCNENGGCDADA 1585
DB 1 NISQHCVKKQCPQNSGCFRHLDERECKLLNKKQSGDKCVENPNTCNENGGCDADA 60
QY 1586 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSN 1621
DB 61 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSN 96
RESULT 22
AAW22593
ID AAW22593 standard; Protein; 127 AA.
XX AAW22593;
XX 25-MAR-1998 (first entry)
XX PFMSPl(p19)S protein sequence.
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX Key Location/Qualifiers
XX

```

```

FT Peptide 1..19
FT /note= "signal peptide"
FT Protein 20..127
FT /note= "mature protein"
FT Region 1..32
FT /note= "derived from P. vivax MSP1"
FT Region 33..34
FT /note= "encoded by restriction enzyme sequence used to
FT create the chimeric sequence"
FT Region 35..127
FT /note= "derived from P. falciparum C-terminal p19
FT fragment of MSP1"
XX WO9730159-A2.
XX 21-AUG-1997.
XX 14-FEB-1997; 97WO-FR00291.
XX 14-FEB-1996; 96FR-0001821.
XX (INSP ) INST PASTEUR.
XX (UYN ) UNIV NEW YORK STATE.
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX Roth C;
XX WPI; 1997-425034/39.
XX P-PSDB; AAW22592.
XX Recombinant protein containing plasmodium merozoite surface
PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT antibodies for diagnosis and protein purification
XX Disclosure; Fig 1C; 85pp; French.
XX This is the amino acid sequence of a chimeric protein comprising amino
CC acids 1-32 of the plasmodium vivax merozoite surface protein 1 (MSP1)
CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX Sequence 127 AA;
XX Query Match 6.4%; Score 539; DB 18; Length 127;
XX Best Local Similarity 98.9%; Pred. No. 8.6e-17;
XX Matches 92; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1526 NISQHCVKKQCPQNSGCFRHLDERECKLLNKKQSGDKCVENPNTCNENGGCDADA 1585
DB 35 NISQHCVKKQCPQNSGCFRHLDERECKLLNKKQSGDKCVENPNTCNENGGCDADA 94
QY 1586 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCS 1618
DB 95 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCS 127
RESULT 23
AAW36102
ID AAW36102 standard; Protein; 127 AA.
XX AAW36102;
XX 25-MAR-1998 (first entry)
XX PFMSPl(p19)S protein sequence.
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX Key

```

OS Chimeric - Plasmodium vivax.
 OS Chimeric - Plasmodium falciparum.
 XX Key Location/Qualifiers
 FT Peptide 1..19
 FT Protein /note= "signal peptide"
 FT Protein /note= "mature protein"
 FT Region 1..32
 FT Region /note= "derived from P. vivax MSP1"
 FT Region 33..34
 FT Region /note= "encoded by restriction enzyme sequence used to
 FT Region /note= "create the chimeric sequence"
 FT Region 35..127
 FT Region /note= "derived from P. falciparum C-terminal p19
 FT fragment of MSP1"
 XX WO9730158-A2.
 XX 21-AUG-1997.
 XX 14-FEB-1997; 97WO-FR00290.
 XX 14-FEB-1996; 96FR-0001822.
 XX (INSP) INST PASTEUR.
 PA (UYNP) UNIV NEW YORK STATE.
 XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
 PI Roth C;
 XX WPI; 1997-425033/39.
 DR N-PSDB; AAT94549.
 XX Recombinant protein containing the merozoite surface protein-1 p19
 PT fragment - useful in anti-malarial vaccines, diagnosis and protein
 PT purification
 XX Disclosure; Fig 1C; 85pp; French.
 XX This is the amino acid sequence of a chimeric protein comprising amino
 CC acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
 CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kd
 CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
 CC p19 is the C-terminal fragment of the 42 kd MSP1 from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX
 SQ Sequence 127 AA;
 Query Match 6.4%; Score 539; DB 18; Length 127;
 Best Local Similarity 98.9%; Pred. No. 8.6e-17;
 Matches 92; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1526 NISQHCYKQCPQNSGCFRHLDERECKCLLYNQEGDKCVENPPTCENNNGGCDADA 1585
 DB 35 NISQHCYKQCPQNSGCFRHLDERECKCLLYNQEGDKCVENPPTCENNNGGCDADA 94
 QY 1586 KCTEDSGSGKKITCECTKPDSPYPLFDGIFCS 1618
 DB 95 KCTEDSGSGKKITCECTKPDSPYPLFDGIFCS 127
 RESULT 24
 AAR07503
 ID AAR07503 standard; protein; 1254 AA.
 XX
 AC AAR07503;
 XX
 DT 06-FEB-1991 (first entry)
 XX
 DE Merozoite apical-end-localised protein (MAEP) insert 5.3.

XX Malaria; vaccine; Duffy blood group.
 KW Plasmodium vivax.
 OS
 XX WO9011772-A.
 PN 18-OCT-1990.
 PD 03-APR-1990; 90WO-US01849.
 PF 06-APR-1989; 89US-0334270.
 XX 05-APR-1989; 89US-0334041.
 PR (UYNP-) NEW YORK UNIV.
 PA Barnwell JW, Galinski MR, Wertheimer SP;
 XX WPI; 1990-334616/44.
 DR N-PSDB; AAQ06328.
 XX Malarial apical end merozoite proteins and peptide(s) - used for
 PT developing cpds. for treating, preventing and diagnosing malarial
 PT infection
 XX Claim 2; Fig 1a; 66pp; English.
 PS A MAEP compound having a binding affinity for a Duffy blood group
 XX antigen of primate red blood cells, is antigenic for the complete
 CC protein, and may be used in diagnosis, treatment and vaccination
 CC against invasion by P.vivax and P.knowlesi.
 CC
 XX Sequence 1254 AA;
 Query Match 5.4%; Score 455.5; DB 11; Length 1254;
 Best Local Similarity 20.5%; Pred. No. 6.1e-12;
 Matches 319; Conservative 221; Mismatches 469; Indels 548; Gaps 71;
 QY 115 DNSSDSAKSYADLKHVRNVLITIKEL-----KYPOLFDLNFHMLT 156
 DB 8 DTSFDEKKKSTEKAYEKMGN---TLKELEKMDDEKNIKEVEEAOIQKRIF--IDHDVN 62
 QY 157 LC-DNIHGFXYLIDGYE-----EINELLYKLNFPDOLLRAKLVNDVCANDYQCIPFNKIRA 211
 DB 63 LMNDEVEKSKVMKIELYKKEIDKIKQKTNEY-----KQGGDT-SNEY----- 104
 QY 212 NELDVLKLVFGYKPLDNKIDNVGKMEDIY-----KKNKTTIENINELIEESKTKIDK 265
 DB 105 -----YTEQYNSATQSKAKIEQFINIATTKGTSDTSQDINEL-ESIKEEVHK 151
 QY 266 NKNATKE-----EEKKLYOQYDLSIYNKO---LEEAHNLISVLEKRIDTLKKNENIK 316
 DB 152 NLQLVQESNSMEEMRKQILSKWDLILLNSETIAKEISNTQNALGFRENAKTKLNKTD 211
 QY 317 ELLDKINEIKPPPPANGSNTENTLIDKN-----KKIEHEKEIKEIAKTIFNIDSL--F 369
 DB 212 ELLQVAAIMEEAKRAHKNIDIALEDAQIDTEVSKIEQINREIMNKDKDEIKSYLSEIREY 271
 QY 370 TDPLELYLREKKNKIDISAKVETKESTEPNEYPNGVTYPLSYNDINNALNELSFGDL 429
 DB 272 KDKCTTEISNSKRGKD-----KIEFLEKFKPNEESNS-----NKNV--INEINE--NI 315
 QY 430 INPFDTYKE-----PSKNITDNE-----RKKFIN---EIKEK 459
 DB 316 RNSEQYLDIEDAEKQASTKVELPHKHTTISNIFKSEILGVETKSOKKINKAEDIMKE 375
 QY 460 IKIEKKKTESDKSYEDRSKSLNDITTKYEKLLNFIYDSKFN-----NI- 504
 DB 376 IERNHSEIQTQVKGQENLNKLINE-PHNYDAEDELNDKSTNAKVLIETNLESVKHNL 434
 QY 505 DLTFNEKMMGKRYSYKVE-----KLTHNT-----FASYEN----- 535
 DB 435 EITNIKOGGEKIYSKAKDIMOKIKATSENTAEKTLKDYKDDOSNVYVNTTTERNLIV 494


```
QY 536 -----SKHN-----LEKL-----TKALK----- 548
DB 495 TEKNRLNGIDSTITNIEGALKRESKNGYEIGFLEKLEEGKRNKLVKVDITKKSINSVTGVNF 554
QY 549 --YMEDYSLRNIVBEKELKYKNLSKTENEIE-----TLVENIKDEBOLFEKTKTKDEN 602
DB 555 SSLFNFNOLQYDFNKNINDYENKNGEYNEFEGLNRISENLRNASEN-----TSDYN 608
QY 603 KPDEKILEVSDIVKQVQKVLMMKNKIDELKKTQLTLKKNVELKHNHVPNSVKQENKQEPY 662
DB 609 SAKTURL-----AQEKVNLNKEEANKYLRDVKVE-----SPR----- 645
QY 663 YLIVLKEIDKLKVPMPKVESLINEEKNKIK-----TEQSDNS 701
DB 646 FIFNMKESLDKINEMIKKEQLTVNGBGNVQOLVENIKELVDENNLSDILKAQATGKNEEI 705
QY 702 EPSTGEITGOATKFGQOAGSA-----LEGDSVQAQAQOQKQOQPPVPV 746
DB 706 QKITHSTLKNKAKTILGHVDTSAKYVGIGITPELALTELLGDAKLTAQELKF----- 758
QY 747 PVPEAKAQVPTPPAPVNN---KTENVSKLDVLEKLYEFLNTSYICHKYILVSHSTMNEKI 803
DB 759 ---ESK-----NNVLETKNSK-----TNEL-----DVHKN 784
QY 804 LKQYKITEESKSSCDPLDLLEFNIONNIPVMYSMFDSLNNLSOLFMEIYKEMVCNL 863
DB 785 QDAYKVALE---ILAHSEIDT-----KOKDS---SKLIEMGNQIYLVKVLINQ 827
QY 864 YKLKDNKIKNLEBAKVSYSVTKLSSSSMOPLSLTPQDKPEVSANDDTSHSTNLNLSL 923
DB 828 YK---NKISSIKKEEAVSVKIGNVSKKHSKLSKITSKDSK----- 865
QY 924 KLFENILSIGKN---KNYIQELIGOKSSENFYEKTLKDSDFYNESFTNFVSKRADDINS 980
DB 866 --YDNIILEKQTELQNLNRNSTOEXTNTNSDSKLEKIKTFD---ESLKNALKTLEGEVNA 921
QY 981 L-----NDESKKKLEEDINKLTKTLQSLFDLYNKKYIKLRLERLDFKKT---VGRYKMOIK 1033
DB 922 LKASSDNHEHVQSK-SEVPNPALSEIEKEETDIDSLNTALDELLKGRTECVSRYK----- 976
QY 1034 KLTLLKEQLESKLSLNNPKHVLFQNFVFFNKKKEAEAEFTENTLE-NTKLILKHYKGLV 1092
DB 977 ---LIKOTV-----TKEISDDTEILNTIEKNVKAYLAYIK--- 1008
QY 1093 KYINGESSPLKLTSESIQTEDNYASLENFKVLKLECKLDNKLNLKLSYLSLSSGLHH 1152
DB 1009 KNYEDTVQDVLTLNEHF-----NTQVSNHEPTNFDKSNKSSSELTKAVIDSKT 1057
QY 1153 LIAELKEVIKNKNYTGNSPSENNTDVNNALSYKFKFLPEGTDVATVWSEGSSTLEQSQP 1212
DB 1058 IISKLGVIIEVN-----ENTEM-NTIESSAK-----EIEALYNE-----LKN 1094
QY 1213 KKPASTHVGAESNITTTSQ---NVUDEVDVVIPIFGESEEDYDLDGQVVTGEAVTPSV 1269
DB 1095 KKTSLNEIYQTSNEVKYKLEQMSNADKYID----- 1123
QY 1270 IDNLSKITEYEVLYLKLPLAGVYRSKLENNVMTFNVVKNVDTLNSRENKFNKVL 1329
DB 1124 ---VSKIFNT-----VLDTQKSNIVTNOHSIN-NVKDKLKGK-----LQELI 1161
QY 1330 ESDLIPYKDLTSSVNVKDPYKFLNKKRDKFLSSYNYIKDSIDTDINFANDVLGYKIL 1389
DB 1162 DAD-----SSFTL-----ESIKKEFENEIYSHIKINI-----GELBQL 1192
QY 1390 SEKYSKLDLSTKYYKINDKQGENEYKLPPLNNIETLYKTV-NDKIDLFIHLEAKVLN 1445
DB 1193 QOTNKSSEHNDYAKH-----KEKIVHLINRVESLKGDKVKNHDDQYMKKLNASILN 1242

RESULT 25
AAW24575
ID AAW24575 standard; Protein; 1254 AA.
```

```
XX
AC
AC AAW24575;
DT 10-NOV-1997 (first entry)
DE Merozoite apical-end protein clone 5.3.
KW Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;
KW human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;
KW Duffy blood group antigen; red blood cell; therapy.
OS Plasmodium vivax.
PN US5646247-A.
PD 08-JUL-1997.
XX 05-APR-1989; 89US-0334041.
XX 04-OCT-1991; 91US-0792865.
XX 05-APR-1989; 89US-0334041.
PR 06-APR-1989; 89US-0334270.
PR 03-APR-1990; 90WO-US01849.
PR 02-NOV-1990; 90US-0608639.
XX (UJNY ) UNIV NEW YORK STATE.
PA Barnwell JW, Galinski MR;
PI WPI; 1997-362995/33.
DR N-PSDB; AAT80072.
XX Plasmodium merozoite apical end protein - useful as antigen for
production of anti-malarial vaccines
XX Claim 1; Column 29-38; 68pp; English.
XX AAW24575 and AAW24576 represent the merozoite apical end proteins (MAEP)
isolated from two different Plasmodium vivax strains. These proteins are
the antigens of the invention, and immunoreact with antibodies against a
native MAEP sequence. P. vivax is one of the four malarial species that
infects humans, and is difficult to target for a vaccine, as it cannot
be cultured in vitro. The preinvasion orientation of malarial merozoites
indicates that the apical end plays an important role in the invasion
process. The MAEP protein binds to the surface of susceptible
erythrocytes from P. vivax susceptible humans and primates, and also
binds to rabbit erythrocytes. The antigen can be used for the production
of antimalarial vaccines. The antigens are involved in the invasion
process, and are immunochemically reactive with antibodies raised
against malaria (particularly P. vivax) blood stage parasites. Synthetic
proteins, polypeptides, peptide fragments and analogues of these
antigens can be used similarly. As the antigens specifically bind to a
Duffy blood group antigen (the antigen present on the surface of
susceptible mammalian red blood cells), and are necessary in the process
of invasion of red blood cells by merozoites, they can be used to inhibit
the invasion of red blood cells by a malarial organism. The antigens can
also be used in a method for inhibiting invasion of susceptible
mammalian blood cells by malarial merozoites, and in a method for
inhibiting the propagation of a malarial organism in susceptible red
blood cells.
XX
SQ Sequence 1254 AA;
Query Match 5.4%; Score 455.5; DB 18; Length 1254;
Best Local Similarity 20.5%; Pred. No. 6.1e-12;
Matches 319; Conservative 221; Mismatches 469; Indels 548; Gaps 71;
QY 115 DNSSDSDAKSVADLKHVRVNYLLTIKEL-----KYPQLFDLTNHLMT 156
DB 8 DTSEDEKKKSIEKAYEKMGN---TLKELEKMDDEKNIKEVEEAAIQYKRIF--IDHVDN 62
QY 157 LC-DNIHGFYKYLIDGYE----EINELLYKLNFYDLLRAKLNDCANDYCOIPFNLKIRA 211
```

Db 63 LMNDEVEKSIWKEIYKKEIDKOKTNEY-----KQGD-T-SNFY----- 104
Qy 212 NELDLVKLVFGYRPLDNIDKONVGMEDYI-----KKNKKTITENINELIBESKTTIDK 265
Db 105 -----YEQYNSATOSKAKIEQFINIATTKGTSDTSODINEL-ESIEEVEHK 151
Qy 266 KKNATKE-----EKKKLYQAQYDSIYNKQ-----LEAHNLISVLEKRIDPLKKNENIK 316
Db 152 NLQLVKQESNMEERKQILSKMDLLILNNSETIAKEISNTQNALGFRENAKTKLNKTD 211
Qy 317 ELLDKINEKKNPPANSNPFTLIDKN-----KKIEHEKEIKEIAKTIFKNIDSL--F 369
Db 212 ELLQRAAMIEBAKAKNNIDIALEDAQIDTEVSIEQINRMKQDEIKSYLSEIKY 271
Qy 370 TDPLEYVLEKKNIDISAKVETKESTPEPNEYPNGVTYPLSYNDINNALNELSFGDL 429
Db 272 KQKTEITSNRGKD-----KIEFLEKFPNEENS-----NKVN--INEINE--NI 315
Qy 430 INPFOYTKR-----PSKNIYTDNE-----RKFFIN--EIREK 459
Db 316 RNSEQYLDIEDAEKQASTKVELFHKHETTISNIFKSEILGVETKSKKINKAEDIMKE 375
Qy 460 IKIEKKKTESDKSDESKSLNDITKEYEKLLNELYDSKFN-----NI- 504
Db 376 IERHNSIEQTVKGFQENLKNLNE--PHNYDNAEDELNDKSTNAKVLITETNLESVKHLS 434
Qy 505 DLTNPEKMMGKRYSYKVE-----KLTHHT-----PASVEN----- 535
Db 435 EITNIKQGEKIYKAKDIMQIKATSENTAEKTLKVKDDQSNVYVNLQITTERNLIV 494
Qy 536 -----SKHN-----LEKL-----TKALK----- 548
Db 495 TEKNRLNGIDSTITNEGALKESKNGYEIGFLEKLEIGCKNRKLVITKKSINSTVGNF 554
Qy 549 ---YMEDYSRLNIVWEKELYKNLISKTEBETE-----TLVENIKKDEQOLFPEKKTIDBN 602
Db 555 SSLFNFDLQNYDFNKNIDYENKGEIYNEFEGLSKISENLNASEN-----TSYN 608
Qy 603 KPDEKILEVSDIVKQVQKVLMMNKIDELKQTOLILKNVLEKHNIHVPSYKQENKQEPY 662
Db 609 SAKTLRLE-----AQEKVNLLNKEEAEANKYLRDVKVE-----SFR----- 645
Qy 663 YLIVLKKEIDKLVKMPKVESLINEKKNIK-----TEGOSDMS 701
Db 646 FIFNMKESLDKINEMIKKQOLVNEGHVGNVKNOLVENIKELVDENNLSLILKQATGKNEI 705
Qy 702 EPSTEIGITGOATTRPGQOAGSA-----LEGDSVQAOAQEQKQAPPPVP 746
Db 706 QKITHSTLKNKAKTILGHVDTSKAYVGIKITPELATTELLGDAKLKTAQELKF----- 758
Qy 747 PYPEAKAQVPTPPAPVNN---KTENVSKLDYLEKLYEFLNTSYICHKYLVSHTMNEKI 803
Db 759 ---ESK-----NNVLETENMSK-----NTNEL-----DVHKNI 784
Qy 804 LKQYKITKEESKLSGCDPLDLFIQNNIPWYSNFOSLNSLSOLFMEIYEKEMVCLN 863
Db 785 QDAYKVALE---ILAHSDIEDT-----KOKDS--SKLIEMGNQIYLVKVLINQ 827
Qy 864 YKLKNDKIKNLLEAKKYSTSVKTLSSSSMQPLSLTQDKPEVSAANDTSHSTLNLSL 923
Db 828 YK-----NKISSIKSEEAHSVIGNVSKHSHSLSKITCDKS----- 865
Qy 924 KLFENILSLGKN---KNIQOELIGOKSSSENFYKILKDSDTFYNESFTNVFKSKADDINS 980
Db 866 ---YDNIIALEKOTELONLNSFTQKTNWSDSKLEIKITDP--ESLKNALKTLEGEVNA 921
Qy 981 L-----NDESKRKKLEEDINKLTKTLOISFDLYNNYKYLKLERLFDKKT--VGKYMOTK 1033
Db 922 LKASSDNHERVQSK--SEPNVPALSETEKEBETDIDSLNTALDELLKKGRTCEVSRYK---- 976
Qy 1034 KLTLLEQLESLNSLNNPKPHVLONFVSFFNKKKAEATAETENTLE--WTKILLKHVKGLV 1092
Db 977 ----LIKDTV-----TKEISDDTELTINTIEKNVKAIVAYIK---- 1008

RESULT 26

AAB18172
ID AAB18172 standard; Protein; 2485 AA.
XX
AC AAB18172;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:29.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
XX WPI; 2000-365347/31.
XX
DR
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite.
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 75-82; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are refined or secreted or membrane proteins)

Staphylococcus epidermidis SRI strain; infection; diagnosis;
vaccination; endocarditis.

Staphylococcus epidermidis.

WO200134809-A2.

17-MAY-2001.

09-NOV-2000; 2000WO-US30782.

09-NOV-1999; 99US-0164258.

(GLAX) GLAXO GROUP LTD.

Kimmerly WJ;

WPI; 2001-316495/33.

N-PSDB; AAH53785.

Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
useful for vaccinating against infections, e.g. endocarditis -

Claim 18; Page 779-781; 2188pp: English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the S. epidermidis polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to AAH55090 represent specifically claimed S. epidermidis genomic DNA polynucleotide sequences from the present invention. AAH55091 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention.

N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though the sequences are given in the disclosure for SEQ ID NO:4472, no sequences are present for SEQ ID NO:4455 to 4464.

Seq Sequence 5024 AA;

Query Match 5.3%; Score 450; DB 22; Length 5024;

Best Local Similarity 20.6%; Pred. No. 5.6e-11;

Matches 360; Conservative 268; Mismatches 653; Indels 466; Gaps 84;

QY 107 NSRRTPNSDSSDAKSYADLKHVRNVLITIKELKYPOLF--DLTNHMLTLCN----160

Db 3011 NKAKALNDMMKKLDIVAQQDNVRGNNF---INEDSTPQNNYNDTINHAQSIIDOVAMP 3067

QY 161 -----IHGFYKLI---DGYEINELLYKLFYDLIRAKLNDVCANDYCQIPFNL 207

Db 3068 TMSHDEIENAINNHKAINALDGEHKLQAKENAN---LLNSLND-----3110

QY 208 KTRANELDLVKLV--FGYRKPLDNKDNVGMEDYIKKNKTTININELIEESKKTIDK 265

Db 3111 -LNAFORDAINRLVNEAOTREKVAEQLOSAQALNDAMHLRNSIQNSSVRQESKYI---3166

QY 266 NKNATKEEEKKLYQAOYDLSYLNKOLEAHNLISVLEKRIDTFLKKNENIKELDKINEI 325

Db 3167 NASDARKEO-----YNHAVREVENIINEQHPTLD-----KEIKQLTDAVNO- 3208

QY 326 KNPPANSNGTNTLLDNKKTKEE-----HEKETKEIAKTIKEN-----IDSLFTD 371

Db 3209 -----ANNLDNGVELLDADQNAHOSIPTMLHNLQAQNALNEKINNATRAKVAALICQ 3263

QY 372 PLELEYLREKNNKIDISAKVETKEST-----EPN---EYPNGVTY-----PLS 412

Db 3264 AKILDHAM--ENLEESIKDKQOVKQSSNYINEDPDQVQETYNNAVDHVTIELNQTNPNTLS 3321

QY 413 YNDINNALNELNSF-----GDLINPDYDT--KEPSK-NIYTDNERKKFINET--REKIKI 462

Db 3322 IEDIEHAINEVNAQKQJRGKOKLYQTIDLADKEKLSLDLTSQOSSISNOLYTAKTAT 3381

QY 463 EKKKTESDKSYEDRSKSLNDITKEYEKLNLNIYDSKFNNNIDLTNFEKMMGKRYSYKVE 522

Db 3382 EVAQAIEKAKSLNHAMKALNKIYKNADKVID---SSRFINE---DOPEKEAQQAINHVD 3435

QY 523 KLTHHT-----FASYENSKHNLEKLTALKYMEDYSLRNI-----V 559

Db 3436 STIHRQTNPEMDPTVINSTHETEAQNNLHGDQLAHAKQ-----DAANVINGLIHLNV 3490

QY 560 VEKEL-----KYKKNL--ISKIENEIETLVE-----NIKKDEOLFEEKKIK 599

Db 3491 AOREVMINTNTNATPREKAKNLDNAQALDKAMETLQQVVAHKHNLNLDNLSKYLNED--SK 3548

QY 600 DENKPEKILEVSDIVKVOQVQVLLMKNKIDELKKTQL-----IL-----KNVEL 643

Db 3549 YQQQYDRVIADEAQLLNTTNPTEPYKVDIVKDNVLANEKILFGAEKLSYDSKNADEI 3608

QY 644 KNIHVPNSYKQENKQ-----EPYLIYLVKKEIDK-LKVFMPKVESLINEEKKNIK 693

Db 3609 KHMNLLNNAQKOSIKDMISHAALRTEVKQLLQQAQKTLDEAMKSLEDKTOVWITDTLPNY 3668

QY 694 TEGQSDNSEPSPEGETGQATTKPGQAGSALGEGSVQAOEQKQAOQPPV--PVPVPEA 751

Db 3669 TEASEDKKEKYDQTVSHAQAIIT--DKINGSNVSLQVROALQQLQASENLQDQVBEA 3726

QY 752 KAQVPTPAPVNNKNTENSKLDYLEKLYEFLNTSYICHKYILVSHSTMNEKILKQYKIK 811

Db 3727 KV-----HANQITDQLTHLSLQO-----QTAKESV---KNATK 3757

QY 812 EESKLSGCDPLDLLENQNNIPVMYSMPEDSLNLSQLFMEIYKEMVCNLYKLKDNK 871

Db 3758 LEEIATASNAL-ALNKVMGKLEQFINHADSIENS-----DNYRQADDDK 3801

QY 872 I---KNLLEAKKV--STSVKTLSSSQMPLSLTPQDKPEVSANDDTSHSTNLSL-KL 925

Db 3802 ITAYDDALEHGDIOKSNATQNEAKAQLQO-----INAETSLNGFERLNHARPA 3852

QY 926 FENILSLGKNKIYQELIQKSSENF---YEKILKDSOT-----FYNESTFNVKSK 974

Db 3853 LEYIKSLEKINNAQKSALEDKVTQSHDLLEHLVNEGTLNMDIMGELANAIVNNYATPK 3912

QY 975 ADDINSLNDESKRK-KLEEDINKLKTLL-----QLSFDLYNKYKLKLERLFDKKKTVGK 1027

Db 3913 A-SIANYINADNLJRKNFTQAINNARDALNKTOGQNLDFNAIDTEK---DDIFTRKDALN- 3967

QY 1028 YKMQIKKLTLLKEQLESKLNLSNPKHVLQNFVSFFNKKK---BAEIAETENTLENTKI 1083

Db 3968 ---GIERLTAASKAEKLIDSLK-----FINKAQFTHANDEINTNSIAQLSRI 4013

QY 1084 LLKHV-----KGLVYKYNNGESSPLKTL-----SEESIOTEDNYASLENKVLKLEGL 1132

Db 4014 VNOAFDLNDAMKSLRDELNNQAPVQASSNYINSDLDLQKQFDHALSNARKVLAKENGKN 4073

QY 1133 KDNLEKKKKLSYLSGSLHLLIAELKEV-----IKKNYTCNSP---S 1172

Db 4074 LDEIQIE-----GLKQVIEDTKDALNGTORLSKAKAKAIQVQSLSYINDAQRIHA 4124

QY 1173 ENNTDVNNALESYKKFLPEGTQVATVVSBSGSDTLQSQPKKPASTHVAESNTITTSN 1232

Db 4125 ESNINSDLSLANTLWSKASDLNAMKDL-RDTLESN-----STSVNSVNYINADKN 4177

QY 1233 VDEYDDVI-----IVPIFGESEEDYDGLQGVVTGEAVTPSVIDNLSK 1276

Db 4178 LQIEFDEALQOASATSSKTSNPATIEEVLGLSQAIY-DTKNALNGE-----OR 4225

QY 1277 IENEYEVLYLKLPLAGYRSLLKQLENNVMTFNVNVDILNSRENKFNKNVLESPLIY 1336

Db 4226 LATE-KSKDLKLIKGLKDLNKAQLE-----DVTN-----KVNSANTLFE-----L 4264

Qy 1337 KDLTSSNYVVRDPYKFLNKKERDRKFLSSVYIKDSI---DFTDFINFANDVLGYKILSKY 1393

Db 4265 SOLQSTLKLNDKMKLL---RDLKTLVNPVKASLNVNRADYN-----LKRQF 4309

Qy 1394 KSDLDSTIKKYINDKOGENEKYLPLFNNTETLYKTVD-----KIDLFVI 1437

Db 4310 NKALKEAGVGNKNSGTNVN---INDIOHLLTQIDNAKDQDLNGERRLKEHQKSEVFII 4365

Qy 1438 HLEAKVLNVTYE-----KSNVEVKI-----KELN-VLKTODKLADF-----KKN 1477

Db 4366 K-EDILNNAOKAALINOIRASKDIKIINOIVDNAIELNDAMQGLKEHVAQLTATTKDNI 4424

Qy 1478 NFVGT-ADLSYDYNH-----NNLLTKFLSTGMVFENLAKTVLSNLLDGN--LOGMLNISQ 1529

Db 4425 EYLNADEDLKIQDYAINLANNVLDKENGTKN-DANIIIGMIQNMDDARALLNGIERLKD 4483

Qy 1530 HOCVKKQCPQNSGCFRHLDEREECKLLNYKQEGDKCV-ENPNPTCNENNGCDAKCT 1588

Db 4484 AQ-TKAHNDIKDTLKRQDLETEHANATSNKSAQAKQVMYNEBARKAFSINHATSNDLVNQ 4542

Qy 1589 EEDSGSN 1595

Db 4543 AKDEGQS 4549

RESULT 28

ABP38314

ID ABP38314 standard; Protein; 10182 AA.

XX AC ABP38314;

XX 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

KW antibacterial; gene therapy.

XX Staphylococcus epidermidis.

OS Staphylococcus epidermidis.

XX USG380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

DR N-PSDB; ABN90859.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis

PT polypeptide, useful for diagnosing and treating bacterial infections -

XX Disclosure; SEQ ID 3159; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading

CC frame (ORF) nucleic acid sequences which encode the amino acid sequences

CC given in ABP35124 to ABP37960. The S. epidermidis sequences have

CC antibacterial activity and can be used in gene therapy. The sequences

CC can also be used in the diagnosis and treatment of bacterial infections,

CC particularly S. epidermidis infections. The sequences can be used to

CC screen for compounds able to interfere with the S. epidermidis life

CC cycle or inhibit S. epidermidis infection.

CC N.B. The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC USPTO web site.

XX Sequence 10182 AA.

Qy Query Match 5.3%; Score 448; DB 23; Length 10182;

Best Local Similarity 20.6%; Pred. No. 1.6e-10;

Matches 364; Conservative 250; Mismatches 651; Indels 498; Gaps 81;

Qy 107 NSRRTNPSDSSDAKSYADLKHRVRYLLTIKELKYPQLF--DLTNHMLTLCN---- 160

Db 7957 NKALNNDMKKLKDIVAQODNVROSNNY---INEDSPQNNYNDTINHAQSIIDVAVP 8013

Qy 161 -----IHGRKYLI---DGYEINELLYKLFYDILLRAKLVNDVANDYCOIPENL 207

Db 8014 TMSHDEIENANNIKHAINALDGEHKLQAKENAN-----LLINSND----- 8056

Qy 208 KIRANELDVAKLV--FGYRKPLDNKIDNVGKMDYIKKKKTKIENINELIEESKTTDK 265

Db 8057 -LNAPORDAIIIRLVNEAQTREKVAEQLOSQAALNDAMKHLRNSIQNSVROESKYI--- 8112

Qy 266 KKNATKEBEKKKLYQAQYDLSYINKQLEEAHNLISVLEKRIIDTLKKENIKELLDKINEI 325

Db 8113 NASDAKKEQ-----YNHAVREVENINEQHTFLD---KEIIKQLTDGVNQ- 8154

Qy 326 KNPPANSNGTPTLLDNKKIE-----HEKEIEIAKTIKFNIDSLFTDPL----- 373

Db 8155 -----ANNLNGVELLDADKQNAHQSIPTLMHLNQAQNALNEKINNNAVTRTEVAALIQ 8209

Qy 374 -----ELEYLREKNKNIDISAKVETKESTEPNPGVTY-----PLSY 413

Db 8210 AKLLDHAMENLEESIKDK-EQVKOSSNYINEDSDVQETYNDAVDHVTILNQTVMPTLSI 8268

Qy 414 NDINNALNELNSF-----GDLINPFDT--KEPS-KNIYTDNERKKKINEI-KEKIE 463

Db 8269 EDIEHAINEVNAQAKQLRGKQKLYQTDLADKELSKLDLTSQQSSSSISNQIYAKTTE 8328

Qy 464 KKKIESDKKSYEDSKSLNDITKEYELLNFIYDSKFNNIDLTNFEKMGKRYSYKVEK 523

Db 8329 VAQAIERAKSLNHAMKALNKYKNAQKVLQ---SSRFINE---DQPEKKAYQAQINHVDS 8382

Qy 524 LTHNT-----FASYENSKHLEKLTALKKYMEDYSLENI-----VV 560

Db 8383 ITHROTNPMDPTVINSTHETAEANLHGDKLAHA-----QODAAVINGLIHLNVA 8437

Qy 561 EKEL-----KYKNL--ISKTEIETIVE-----NIKKDEEOLFEEKTKD 600

Db 8438 QREVMINTNTATTREKVAKNLDNAQALDKAMETLQOVVAHKNNILNDSKYLNE--SKY 8495

Qy 601 ENKPDEKILEVSDIVKQVQVLLMKNIDELKKTQL-----IL-----KNVELK 644

Db 8496 QQQYDRVIADAEQLLNQTTNPTLEPKYVDIVKDNVLANEKILFGAEKLSYDKSNADEIK 8555

Qy 645 HNIHVPNSYKQENKO-----EPYLLIVLKKREIDK-LKVFMPKVESLINEEKKNKT 694

Db 8556 HMNYLNNAQKOSIKDMISHAALRTEVKOLQOAKILDEAMKSLEDKTOVITDTPLPNT 8615

Qy 695 EGQSDNSEPSTEGETTGQATTKPGQAGSALEGSVQAQAEQQAQPPV--PVPVPEAK 752

Db 8616 EASEDKKEKVDQTVSHAQAI--DKINGSNVSLQDQVRAQLQTOASENLGDQDQVEAK 8673

Qy 753 AQVPTTPAPVNVKNTENYSKLDYLEKLYEFLNTSYICHKYILVSHSTMKEKILKYKTYKE 812

Db 8674 V-----HANOTIDOLTHLSL-----QOOTAKE 8696

Qy 813 EESKLSGCDPLDLFNIONNIPVMYSMDLSLNSLSQLFMEIYEKEMVCNL--YKLKND 870

Db 8697 ---SVKNATKLEETATVSNA-----OALNKVNGKLEQFINHADSVENSQYROADDD 8746

Qy 871 KI-----KNLEAKKVSTSV---KTLSSSSMOPL-----SLT 900

Db 8747 KIIAYDEALEHGQDIQKTATNQNETKQALQQLIYAETSLNGFERLNHAPRALEYIKSL 8806

Qy 901 PQDKPEVSANDTSHSTNLNLSKLFIENLSLGKKNKIYQELIGQ-----KSS 948

Db 8807 KINNAQSALED---KVTQSHDLLEHIVNEGTLN---DIMGELANAIYNNYAPTRAS 8860
Qy 949 ENFYEKILKSDTFYNESFTNFVSKADDISLN---DESKRKLLEEDINKLKTLOLS 1004
Db 8861 INYI-----NADNLKKDFTQAINARDALNKTQGNLDNFADTFKDDIFKTDALN-G 8914
Qy 1005 FLYNKYKILKLERLFDKKTGKVKYK-----QIKKLT-----LKEQLESKL 1046
Db 8915 IERLTAASKAKELDLKLFINKAQFTHANDEIMNTNSTAQLSRVNQAFDLNDAMKSLR 8974
Qy 1047 NSLNNPKVHVNFSVFFNKKKEAEATAENTLENT-KILLKHVGLVYKNGESSPLKTL 1105
Db 8975 DELNNAQFPVQASSNYINSDEDLK-QQFDHALSNARKVLAK-----NG-----KML 9020
Qy 1106 SESEIQ-----TEDNYASILENFVKLSLEGLKDLNLEKLSYLSGLHHLIAELREV 1160
Db 9021 DEKIQGLKQVETEDFKDALGQIRUSKAKAIAQV-----QSLSYINDAQRI----- 9069
Qy 1161 IKNKYNTGNPSSENNTDVNNALLESYKFLPECTGVATVVSSESGDTLQSOQPKPASTHV 1220
Db 9070 -----AENITHSDOLSLANTLSKASDLDNAMKDL-RTIESN-----STSV 9111
Qy 1221 GAESNTITTSQNVDEVDVI-----IPIFGESEEDYDGLGQVVTGEA 1264
Db 9112 PMSVNYINADKNLQIEFDEALQOASATSKTSENPAETIEEVGLSQAIY-DTKNALNGE- 9169
Qy 1265 VTPSVIDNLSKIENEYEVLYLKLPLAGVYRSLLKOLENNVMTFNVKDLNLSRNFKNREN 1324
Db 9170 -----QRLATE-KSKDLKLKGLKDLNKAQLE-----DVTN-----KVN 9202
Qy 1325 FKNVLESDLIPYKDLTSSNYVVKDYKFLNKEKRDKFLSSNYIKDSI-----DTDINFAND 1381
Db 9203 SANTIETE-----LSQLTOSTLELNDKMKLL-----RDKLTLVNPVKASLNYRNADYN----- 9250
Qy 1382 VLGYKILSEKYSKLDLSIKKYNKOGENKYLPLFLNNIFLYKTVD----- 1430
Db 9251 -----LKQPNKALKEAGVNLKNSGTNVN-----INDIHLTLQIDNAKDQNGERRL 9299
Qy 1431 -----KIDLFVHLBAKLVNYE-----KSNVEVKI-----KELN-----YLKTI 1466
Db 9300 KEHQKSEVFIK-ELDLINNAQKAAIINQIRASKDIKIINQIVDNAIELNDAMQGLREH 9358
Qy 1467 QDKLADFKNNNFVGIADLSTYNH-----NNLLTFLSTGMVFENLAKTVLSN 1515
Db 9359 VAQLTATTKDN-----IEYLNADHDHKLQDYAINLANNVLDKENGNTK-DANIIIGMIQN 9413
Qy 1516 LLDGN--LOGMLNISQHCQVKQCPQNSCCPRHDERECKLLNYKQEGKCV-ENPNP 1572
Db 9414 MDDARALLNGIERLKDAQ-TRAHNDIKDTLKRQLDDEIEHANATSNSKAQAKQMVNEEAR 9472
Qy 1573 TCNENNGCDADAKTEEDSGSN 1595
Db 9473 ALSNINDATSNLDVNOAKDEGOS 9495

RESULT 29
ABG06505
ID ABG06505 standard; Protein; 2633 AA.
XX
AC ABG06505;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6496.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX

PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS70692.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 36864; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 2633 AA;
Query Match 5.3%; Score 445; DB 22; Length 2633;
Best Local Similarity 19.0%; Pred. No. 4.3e-11;
Matches 334; Conservative 327; Mismatches 610; Indels 486; Gaps 78;
Qy 126 ADLKHVRVNYLTITKELK---YPQLFDLTNHHMLT-----LCDNTHGPKY 166
Db 415 AKRKRRTWCLGKINKMKNSYADQFNPTNTTTTKHLSINLLREIDSVCSSESVFSN 474
Qy 167 LIDGYEEI-----NELLYKLNFFDL--LRKLNDCANDYCOI-----PFNLKIR-ANE 213
Db 475 TLDTLSEIENWPATKLLNQENIESSELNLSRADYDNLVL-DYEQLRTEKEEMELKURKND 533
Qy 214 LDV-----LKLVLFG---YRKPLDN-----IKDNVCK 237
Db 534 LDEFALERKTKYKQEMQLIHEISNLKVLKRVYNNQOLENELSKSVKVELLEKEDQIKK 593
Qy 238 MEDYIKKKNKKTIENTIN-----ELIEESK-----TID-KKNKATKEEKKKL 278
Db 594 LQEIYDSQK--LENIKMDLSYSLESTEDPKMQKTLFDAETVALDAKRESAFRLSENLEL 651
Qy 279 YQAQYDLSYNNKQLEAHNLISVLEKRIDTLAKN--ENIKELDKDKINEIKNPPANSNGT 336
Db 652 KEKMKELATTYKOME---NDIOLYOSQLEAKKKMQVDLEKELOSAFNETKITLTSIDGKV 708
Qy 337 PNTLL-----DKNKKIEEHF-----KEIKEIARTIK----- 362
Db 709 PKDLICNLEEGKITDLOKELNKEVEENALREEVILLSELKSLPSEVERLKEIQDKSE 768

Qy 363 -----FNIDSLFDPLELEY-----YLREKNKNIDISAKVETKESTPEPNEYPNGVITYPLS 412
Db 769 ELHIITSEKDKLPSEVVKESRVOGLLEIGTKTKDDLATTQSNVKSSTDQEFQNFKTLHMD 828
Qy 413 YN-----DINNALNEL-----NSFGDLINPFDY-TKPSKNIYTDNERKKF 452
Db 829 FEQYKRVLEENRMRNGEIVNLSKEAQFDSGLGALKTELSYKTOELQEKTRVQERLNE 888
Qy 453 INEIKKIK-----IEKKK-----IESDKKSYEDRSKSLNDITKEEKL 491
Db 889 MEQKLEENRDSPLQVVEREKLITTEKLOOTLEEVATLTOEKDDLKQLOESQIERDQ 948
Qy 492 LNEIYDSKFNINIDLTNFKXMMGRYSYKVEKLTHH-----NTFAS-----YEN 535
Db 949 KSDIHTD-VNMNIDTQ-----EOLRNALLESKQHOQETINTLKSISEEVRNLRHMEEN 1000
Qy 536 ---SKHLEKLTAKLWEDYSLRNI-----VWEKELYKYNLISKIENETL 581
Db 1001 TGTEKDFQOKWVIGDKKQDLAKNTQTLTADVKDNEIIEQQRKIF-SLIQE-KNEIQOM 1058
Qy 582 VENIKKDEEOLFEKKITKDN-----KPDEKILEVSDIVQVQVLLMKNKIDELKKQTOL 636
Db 1059 LESVIAEKEOL---KTOLKENIEMTIENQELRLIGDELKKO-OEIVAQEKNAHAKKEGE 1114
Qy 637 ILKN---VELKHNHPNSYKQENKQEPYLIIVLKKKIDKLVFMPKVESLINEER--- 689
Db 1115 LSRCTDRLAEVEEKLKEKSQOLQEQKQ---LLNVQEMSEMOKKINEIENLKNELKNKE 1171
Qy 690 ---KNIKTE---GQSONSEPSSTGEITGQATTKPGQOAGSALBGD-----SVQAAQ 735
Db 1172 LTLHEMETERLELAQKLNENYEEVKSITKERRKVLKELQKSFETERDHLRGYIRIEATGL 1231
Qy 736 EQKQAPPPVPVPEAKAQQPTPAPVNNKTNVSKLDYLEKLYEFLNTSVICHKYLVS 795
Db 1232 QTKELIAHILHEKHQETIDELRRSVSEKTAQIINTQDLEK-----S 1274
Qy 796 HSTMNEKILKOYKITKEESKLSKCDPLDLFNQNNIPVMSFDSLNNS----- 846
Db 1275 HTKLQEEI---PVLHEEQELLPNVKVSQETQETMNELELLTEQSTTKDSTTLARIEMER 1330
Qy 847 --LSOLFMEIYEKEMVNLKLDN-----DKTNLLEAKKYSTSVKTLSSS 892
Db 1331 LRLNEKFOE---SQBEIKSLTKERDNLTKIEALEVKHQDLKEHIRE-----TLAKIQESQ 1383
Qy 893 SMQPSLSTPDQKPEVSANDOTSHSTNLNLSKLFE-----NILSLGKNKIYQELI 943
Db 1384 SKQESLNMKEK-----DNETTKIVSEMEQFKPKDSALLRIEIMGLSKRL----- 1430
Qy 944 GOKSSENFYKILKDSDTFYN-----ESFTNFVYSKADDINSLNDESKR-----KK 989
Db 1431 --QESHDEMSVAKEDDLQRLQEVLOSQESDLKENIKIEIVAKHLETEEELKVAHCCLKE 1488
Qy 990 LEEDINKLK-----KTLQSLFDLYN-KYKLLERLFDKK-----KTVGKYKQIK 1033
Db 1489 QEETINELRVNLSEKETEISTIQKLEAINDKQNKQTOEIEYEEQNLKQISEVQENVN 1548
Qy 1034 KLTLLKE-----QLESKLSNLNPKHVLQNFVSFFENKKK-----EAEIATEN 1076
Db 1549 ELKQFKEHRAKDSALQIESKMLELTLNRLQESQEEIQLMIKEEMKRVQEAQIERDQ 1608
Qy 1077 TLENTKILLKHGVLKYNGESPLKLTSEESI-QTEDNVASLENFKVLSKLGKLDN 1135
Db 1609 LKENTKEI-----VAKMKESQEKYQFLKMTAVNETQKMCETEHLKEQFETQKLNLEN 1662
Qy 1136 LNEKKKLSYLSGLHLLHIAELKEVINKNYTGNSPENNTVDVNNALSKYKFLPEGTDV 1195
Db 1663 IETENIRLTQI---LHENLEEMRSVTKERD-----DLRSVEETLK---VERDQL 1705
Qy 1196 ATVYSESGSDTLEQSPKPPASTHVGAESNTITTSQNVDDVEDDVIIVPIFGSESEYDD 1255
Db 1706 KENLRITRDLQEKELKVHMLKHEQETIDKLRGIVSEKTNEL-----SNMKOLEH 1760
Qy 1256 LGQVVTGEAVTPSIDNLSKIENYEVLXKPLAGVYRSLKOLENNVMTFNVNVDKIL 1315

Db 1761 SNDALKAQDL-----KIOBELRIAHM-----HLKEQOE-----TID-KLRGIV 1797
Qy 1316 NSRENKRENKVNLESDLIPYKDLTSSNYVYVDPYKFLNKEKRDKFLSSYNYIKDSIDTD 1375
Db 1798 SEKTDKLSNMQ-----KOLENSNAKLOEKIOELKANEHOLI-----TLKKD 1838
Qy 1376 INFANDVLGYKILSEKYKSDLSIKKYINDK-----QGENEKYLPFLN-NIETLYKT 1427
Db 1839 VN-----ETQKVVSEMEQKKQIKDQSLTSLKLEIENLQAEHLHENLEEMKSV 1887
Qy 1428 VNDKIDILFVHLEAKVLNYTEKSNVEVKIKELNY-----LKTIOQKLADFKKNNFVGAD 1484
Db 1888 MKERDNLRRVEETLKLERDQLKESLQETKARDLEITQOELKTCSEKISE-----KTIQISD 1942
Qy 1485 LSTDY--NHNLLTKFLSTGMVFENLAKTVLSNLL-----DGNLOGLMNLISOHCYKQCPQ 1539
Db 1943 IQDLDRKSKDELQKK-----IQELOKKELOLLRVKEDVNMVSHK-KINEMEQLKKOPEP 1994
Qy 1540 NSGCFRHLDERECKCL 1556
Db 1995 NYLCKCEMDNFQTLTKL 2011
RESULT 30
AAM40883
ID AAM40883 standard; Protein; 2688 AA.
XX AAM40883;
XX
DT 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 5814.
DE
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI60039.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5814; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

the encoded polypeptides (AAM38642-AM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed specification.

XX Sequence 2688 AA;

Query Match 5.3%; Score 444.5; DB 22; Length 2688;

Best Local Similarity 18.4%; Pred. NO. 4.6e-11;

Matches 345; Conservative 350; Mismatches 616; Indels 559; Gaps 82;

QY 126 ADLKHRYVNYLLTIKELK---YPQLFDLTNHLMT-----LCDNIHGPKY 166
DB 439 AKRRKRVTWCLGKINKMNSNTADQFNITTTKHLKSLNLLREIDESCSDVFSN 498
QY 167 LIDGVEEI-----NELLYKLNFPDL--LRAKLNDVCANDYCOI-----PFLNKIR-ANE 213
DB 499 TLDTLEIENWPATKLLNOENIESELNSLRADYDNLVL-DYEQLTKEEMELKLEKND 557
QY 214 LDV-----LKKLVFG---YRPLDNL-----IKDNVKG 237
DB 558 LDFEALERKTKKQDEMOLIHISNLKLVKHREYVNDQLENELSKVLELREKEDQIKK 617
QY 238 MEDYIKKKKTTIENIN-----ELIESKK-----TID-KNNKATKEEKKKL 278
DB 618 LQEYIDSKQ--LENTKMDLSYSEIEDPKQKOTFLDAETVALDAKRESAFLRSENLEL 675
QY 279 YQAQDLSLYNKLEEAHNLISVLEKRDITLKKN--ENIKELLDKINEIKNPPANSNT 336
DB 676 KEKMKELATTYKQME--NDIOLYSQLEAKKMKQVDLEKELQSAFNEITKLTSLIDGKV 732
QY 337 PNTLL-----DNKKITEEH-----KEKEIAKTK-----362
DB 733 PKDLLCNLEGGKITDQKELNKEVEENALREEVILLSELKSLPSEVERLKEIQDKSE 792
QY 363 -----FNIDSFTDPLELEY---YLREKKNIDISAKVETKESTPEYNGVTYPLS 412
DB 793 ELHIITSEKDLFSEVWHKESRVQGLLEIGTKDKDATTOSNYKSTDOEFONFKTLHMD 852
QY 413 YN-----DINNALNEL-----NSFGDLINPFQY--TKEPSKNITDNERKKF 452
DB 853 FQKYKMYLEENRMOEIVNLISKEAQFDSSLGALKTSLSYKTQBLQEKTEVEORLNE 912
QY 453 INEIKERIK-----IEKKK-----IESDKSYEDRSKSLNDITKEYKL 491
DB 913 MEQLAEQLNRDPSLOTVEREKTLITEKLOQTLEEVKTLTQEKDDLKQLESQIQRDOL 972
QY 492 LNEIYDSKFNNDITNFKMKMKRYKYVEKLTH-----NTFAS-----YEN 535
DB 973 KSDIHDT-VNMNIDTQ-----EQLRNALESKQHOETINTLKSIISEVSRLNLMHEEN 1024
QY 536 ----SKHNLEKTLKALKYMEDYSLRNI-----VVEKELKYKNLISKIENEIETFL 581
DB 1025 TGCTKDFQOKMVGIDKKQDLKAKKTOITLADVKDNEIIEQORKIF-SLIQE-KNELQOM 1082
QY 582 VENIKKDEQLFEKTKITDEN-----KPEKITLEVSDIVKQVQVLLMKNKIDELKKTQL 636
DB 1083 LGSVTAERQOL---KTDLKENIENTIENQOEURLLLGDELKQO-QEIVAQEKHNAIKKEE 1138
QY 637 ILKN-----VELKHNIHVPSNYQENKQBPYYLIVLKKELDKLVFMPKVESLINEEK --- 689
DB 1139 LSRTCDRLAEVEBEUKLEKSSQLOEKQOQ---LLNVOEEMSEMOKKINEINLKNELKNKE 1195

QY 690 ---KNIKTE---GOSDNSEFSTEGEITGQATTTPGQAGSALSGD-----SVOAQAQ 735
DB 1196 LTLHEMETERLELAQKLNENYEVKSIITKERVKLKELQKSFETERDHLRGYIREATGL 1255
QY 736 EQQAQPPVPVPPAKAQQVTPPPAPVNNKNTENVSKLOYLEKLYEFLNTSYICHKYLVS 795
DB 1256 QTKEELKTAIHLKHEQETIDELRRSVSEKTAQIINTQDLEK-----S 1298
QY 796 HSTMNEKILQYKIFKEESKLSGCDPLDLFNIONNIPVMYSMFDSLNNS-----846
DB 1299 HTKLOEEI---PVLHEQELLPNVKKYSETQETNNELELLTEQSTTKDSTTLAREMER 1354
QY 847 ---LSQLFMEIYECVMCNLYLKDN-----DKIKNLLBEAKKSVTSVTLSSS 892
DB 1355 LRLNEKFOE--SQEBIKSLTKERDNLKTIKALEVHKVDQKHEIRE---TLAKIQESQ 1407
QY 893 SQMPLSLTPQDKPEVSANDDTSHSTNLNLSKLFE-----NILSLGKNKNYQIELI 943
DB 1408 SKQEOSLANKEK-----DNETTKIIVSEMPQFKPKDSALLRIETIEMLGSKRL-----1454
QY 944 GOKSENFEVKILKSDDTFYN-----ESFTNFVKSKADDDINSLNDESKR---KK 989
DB 1455 ---QESHDEMKSVAKKDDDLQRLQEVLOSQDLKENIKIIVAKHLETEBELKVAHCCLKE 1512
QY 990 LEEDINKLK-----KTLQLSFDLYN-KYKILKLERLFDKK-----KTVGKYKMQTK 1033
DB 1513 QBETINELRVNLSEKETEISTIQKOLEAINDKLQNKIQEIEYKEBQLNIIKQISEVQEKVN 1572
QY 1034 KLTLLKE-----QLESKLSLNPNPKVHLQNFVSFFNKKK-----EAEETAETEN 1076
DB 1573 ELKQKPEHRKAKDSALQIESKMKLELTNRLQESQEBEIQIMIKEEMKRVQALQIERDO 1632
QY 1077 TLENTKILLKHYKGVKYVNGESSPLKTLSEESI-OTEDNYASLENKFKVLSLEGLKDN 1135
DB 1633 LKENTKEI-----VAKKESQEKYQFLKWTAVNETQKMCIEIHLKEQFQTKLNLEN 1686
QY 1136 LNLKKKLSYLSGJHJLHIAELKEVIKNNKYTGNSPSENNTDVNNALYESKYKFLPEGTDV 1195
DB 1687 IETENIRLTOI---LHENLEEMRSVTKERD-----DLRSVEETPLK---VERDOL 1729
QY 1196 ATVVSSESQDTEQSQPKKPASTH-----VGAESNTITTSQNVDDVDDV- 1240
DB 1730 KENRETITRDLKQOEELKIVHMHLEKHOETIDKLRGIVSEKTNESNMCKOLEHSNDAL 1789
QY 1241 -----IIVPIFGESEEDYDLQGVVTVGAVTPS-----VIDNLSKIE-----1278
DB 1790 KAQDLKIOEELRIAHMHLEKQOETIDKLRGIVSEKTDKLSNMCKOLENSNAKLOEKIQEL 1849
QY 1279 ---NEYEVLYLK-----PLAGVYRSLKKQLENNVWTF-----NVNVKDIIN-----1316
DB 1850 KANEHQLITLKKDVNETQKVSMEQLAKKQIKDQSLTSLKSLKLEINLNAQKLHENLEEMK 1909
QY 1317 SRFNKRENFKNVLESDLIPYKDLTSSNVVVKDPYFLNKEKRDOKFLSSYNYIKDSITDID 1376
DB 1910 SYMKERDNLRRVEET-----LKERDQOLKES---LQETKARDL 1944
QY 1377 NFANDVLGVYKTLSEKYKSDLSIKKINDKOGENEKYLPPFLNNTIETLYKTVNTDIDFLV 1436
DB 1945 ELQOE-LKTARMLSKHEKTVDKLREKISEKTIQ-----ISDQKOLDKSKDELOKKI 1996
QY 1437 IHLEAKVLNITYEKSNEVYKIKELNYLKTIDQKLADFKN-----NNFVGIADLSTDY 1489
DB 1997 QELQKKELQLLRVKEDVNMHKKINEMEQLKKQ---FEPNYLCKCKMDNFQTKKLHESL 2053
QY 1490 HNNLLTK-----FLSTGMVFENLAKTVLSN-----LLDGNLQGM 1524
DB 2054 EEIRIVAKERDELRRIKESLKMEROQFIAT--LREMIARDRONHVQKPEKRLSDGQOHL 2111
QY 1525 LNIISOHC--VKKOCPONSGCFRHLDEREC-----KCLINY--1559
DB 2112 MESLREKCSRIKELKR---YSEMDHDECLNRLSLDLEKIEFHIMKLVLSYVY 2167

QY 966 SFTNEVKS-----KADDI-----NSLND-----ESRRKKLE 991
DB 1271 HLPNEKSYNINVVVEDISFDIFLISIMDLWETNNNNLLNLINDLLKIYEERKKKIY 1330
QY 992 -----EDINKKTKLQSLPOLYKYLKLERLFD-----KKTUVGKYK----- 1029
DB 1331 ICTSLLLKIFRIKKKSNVFLNIIYAFENDIKLILDSINILIKKVVWTFKNCNDF 1390
QY 1030 -----MQIKKTLLEQELSKLNSLNPKHVLONSVFENKKEAEIAETENTLTK-- 1082
DB 1391 NREKNINIKK--LVKLFFISFYKYLKN--YFLQIYYHFF-----YNNQIYNRKNY 1436
QY 1083 -----ILLKHV--KGLVKVYNGSSPLKTLSESIOTEDNYASLENFKVLSKLEGLK 1133
DB 1437 NFDNFFSFSKYINKIFVEIYSSSSSTSSNSFVFNKSFYMMKMCISIINNMGVVK 1496
QY 1134 DNLNLEK--KLSY-----LSSGLHLHIAELKEVIKKNKYTGNSPSENNTDNNALE 1183
DB 1497 --YINLERVKQVFEHNIMDMVHMKSHLHDI-----DVYGHDSYNNI--YOKIIT 1545
QY 1184 SYKKFLPECTVATVVSSEGSTLEQSQPKPASTHVGASNTIITSQNVDDEVDDVI 1243
DB 1546 SYR--GEEKDLDVIN-----TESVHONRNEDDIDGSINSL 1579
QY 1244 PLFGESEEDYDLGQWVTGEAVTPSVIDNLSKIENEYEVLYKPLAGVYR---SLKKQ 1299
DB 1580 DVFNE-----IMRNIIINHNSLIKDHNDMCTKKKRINFQISSPATSEQ 1623
QY 1300 LENNYMTFNVNVKDIILNSFRKNFNKVNLESDDLIPYKDLTSSNVVVDKPYKFLNKKERD 1359
DB 1624 LMNNHTWNY--LTDVW-----LLOKDYIYNID-----NNMNEHKQN 1658
QY 1360 KFLSSYNIKOSIDPINFANDVLCYKILSE---KYKSDLSLKKYINDKQGENEYKL 1415
DB 1659 VFNKPF-----DNNNNNNNNNNMFLNFYIPENNNNNYRMD---IKKRYPPESYDNNYIM 1711
QY 1416 PFLNNIETLYKTVDKIDLV-----IHLEAKVLNYYEKSVEVKIKEL--NYLKI 1466
DB 1712 --FNNIK-----NEENILLQNNSSSIYIDKLMKDKTEMEPLFNKTKDMKNY--- 1759
QY 1467 QDKLADFKNNFVGIADLSTYDINNLLTK-----FLST-----GMVFENLAKTVLSN 1515
DB 1760 ----NEEQNNELI-----SYPPNNMLQNNIIFVKFFLYTQNLQIIFQNNYIFFLSD 1808
QY 1516 LLDGNLQMLNISOHCV--KKOCPQNSGCFRHLDERECKCLLNYKOGDKCVENPNT 1573
DB 1809 FL-----FINKKEYIEEKKNGQNV-----INIKDEKHITNIK--DGDKHITNIKDG 1856
QY 1574 CENNGGCDADAKTEEDSGSGKKITC-----ECTKPDSPYPLFDGIFCSSSNFL 1623
DB 1857 DKNITNKDDKNITNMKKKNKKNYLTILMYSQECs--FYYSIFNTLI--NDYNEL 1909
RESULT 32
AAAM39097
ID AAM39097 standard; Protein; 2663 AA.
XX
AC AAM39097;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2242.
XX
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
OS
XX Homo sapiens.
XX

PN WO200153312-A1.
PD 26-JUL-2001.
PF 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI58253.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PT Example 4; SEQ ID NO 2242; 10078pp; English.
PS
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX Sequence 2663 AA;
SQ
Query Match 5.2%; Score 441; DB 22; Length 2663;
Best Local Similarity 18.8%; Pred. NO. 6.5e-11;
Matches 323; Conservative 325; Mismatches 578; Indels 494; Gaps 74;
QY 126 ADLKHVRNLYLTIKELK--YPLQFLDTNHLMT-----LCDNIHGFKY 166
DB 415 AKRRRVTCWLGKINKMKNNSYADQFNPTNTTTTKLSINLLREIDSVCSSESVFSN 474
QY 167 LIDGYEEI-----NELLYKLNFFDL--LRAKINDVCANDYQCI-----PFLNKIR-ANE 213
DB 475 TLDTLSEIENWPATKLLNQENIESELNSLRADYDNLVL-DYEQRLTEKKKMKELKEKND 533
QY 214 LDV-----LKLIVFG---YRPLDN-----IKDNVCK 237
DB 534 LDEFEALERKTKKQDMQELIHEISNLKLVKHKREYVYNQDLENLSKVELLEKEKEDQIKK 593
QY 238 MEDYIKKNNKTTIENIN-----ELTEESKK-----TID-KKNKATKEEKKKL 278
DB 594 LOEYIDSKR--LENIKMDLSYSEIETDPKMQKTLFDRAETVALDAKRESAFILRSNLEL 651
QY 279 YQAOYDLSLYNNKQLEEAHNLISVLEKRIIDTLKKN--ENIKELDKINEIKNPPPSNGST 336
DB 652 KEKMKELATYKQME---NDIQLYQSLQEAKKKKMKQVDELEKQSAFNEITKTLTSLDGV 708
QY 337 PNTLLI-----DKNKKIEEHE-----KEIKELATIK----- 362

Db 709 PKDLCNLEGGKTTDQLKLNKEVEENEALREEVILLSELKSPSEVERLKEIQDSE 768
QY 363 -----FNIDSFLDPLEY-----YIREKNKNDISAKVETKESTEPNVPNGVTYPLS 412
Db 769 ELHIITSEKDKLFSEVVKESRVQGLLEIGTKTKDDLATATQSNKSYDQEFQFKTLHMD 828
QY 413 YN-----DINNALNE-----NSGDLINPFY-TKEPSKNITYDNERKKF 452
Db 829 FEQYKRVLEENRMRNOEIVNLSKEAQKFDSSGALKATSELSYQELQEKTRVQERLNE 888
QY 453 INETKEKIK-----IEKK-----IESDKKSYEDRSKSLNDITKYEKL 491
Db 889 MEQKEQLENRDSPLQVREKTLITKEKLOQTLEEVKTLTOEKDDLKQLESQIERDQL 948
QY 492 LNEYDSKFNNDLTNFERKMKRYKYVEKLTH-----NTEAS-----YEN 535
Db 949 KSDIHDY-VNMNIDTQ-----EOLRNALSLKQHOETINTLSKTSIEEVSRLNHEEN 1000
QY 536 ---SKHNLKLTALKYMEDYSURNI-----VVEKEUKYKNLISKIENEIETL 581
Db 1001 TGETKDEFOQKMGVIDKDKQLEAKNTOTLTADVKNDBIIBQQRKIF-SLIQE-KNELQQM 1058
QY 582 VENIKDEEQLFEKITKDEN-----KPDEKILEVSDIVKVOVKVLLMKNKIDELKKTOL 636
Db 1059 LESVIAKEQL---KTDLKENIEMTIENQBELRLGDLKKQ-OEIVAQEKNHAIKKEGE 1114
QY 637 ILKN---VELKNIHVPNSYKQENQEPYILVKKKEIDKLVKFMKPVESLINEEK--- 689
Db 1115 LSRTCORLAEVEEKLEKESQOLQEQO---LLNQVEEMSEMQKKINEIENLAKNELNKE 1171
QY 690 ---KNKTE---GQSDNSPSTGEITGQATTKPGQAGSALGEGD-----SVQAAQ 735
Db 1172 LTLHEMETERLELAQKLNENYEVKSTKTKERKVLKELQKSFETERDHLRGYIREIATGL 1231
QY 736 EOKAQAPVPVPPEAKAQPTPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHKYLVS 795
Db 1232 QTEELKIAHILKEHQETIDELRSVSEKTAQIINTQDLEK-----S 1274
QY 796 HSTMNEKILKOYKITREESKSLSCDDPLLLFIQNNIPVYMSFDSLNNs----- 846
Db 1275 HTKLQEEI---PVLHEEQELLPNVKVSETQETMNELELLTBOSTTKDSTTLARIEMER 1330
QY 847 --LSQLFMEIYEMVKNLYKLDN-----DKTKNLEAKKYSTSVKTLSSS 892
Db 1331 LRLNEKFOE--SOEIKSLTKERDNLKTIKEALEVKHDLQKEHIRE-----TLAKIOESQ 1383
QY 893 SMQPLSLTPQDKPEVSANDTSHSTNLNLSKLFE-----NILSLGKNKIYQELI 943
Db 1384 SKOEQSLNMKEK-----DNETTKIVSEMEQFKPKDSALLRIETIEMGLSKRL----- 1430
QY 944 GOKSSNFYKILKDSFTFYN-----ESFTNFVKSKADDINSLNDESKR-----KK 989
Db 1431 --QESHDEKMSVAKEDDQLRLOEVQLESQDQLENIKEIVAKHLETEBELKVAHCCLKE 1488
QY 990 LEEDINKLK-----KTLQSLFDLYN-KYKLLERLFDKK-----KTVGKYRMQIK 1033
Db 1489 QETINELRVNLSEKETEISTIQKLEAINDKLQNKIQEITYEKEEQINIKQISEVQEKVN 1548
QY 1034 KTLTKE-----QLESKLNLSNPKHVLQNFVSFVNKKK-----EAEIATEN 1076
Db 1549 ELQOFKEHRAKADSALQSTESKMLLELTNRLQESQEEIQIMIKEKEMKRVQEAQIERDQ 1608
QY 1077 TLENTKILLKHVGLVYKNGESSPLKTLSEESI-OTEDNYASLENPKVLKLEGLKLDN 1135
Db 1609 LAENTEI-----VAKMRESQKEYQFLKMTAVNTEQKMCIEHLKQFQETQKUNLEN 1662
QY 1136 LNLKXKLSYLSGLHLHIAELKEVINKNKNYTGNSPSENNTDVNNALESYKFLPECTDV 1195
Db 1663 IETENIRLQI---LHENLEEMSVTKERD-----DLRSVEETLK---VERDQL 1705
QY 1196 ATVSSGSDTLBQOPKPKPASTH-----VGAESNTTITTSQNVDEVDV- 1240

Db 1706 KENLRITITRDEKQEEELKIVHMLHKEHQETIDKLRGIVSEKTNISNMOKDLEHSNDAL 1765
QY 1241 -----IIVPIFGESEEDYDDLGOVVTGEAVTPS-----VIDNILSKIE----- 1278
Db 1766 KAODLKTOEEELRTAHMLHKEQOETIDKLRGIVSEKTDKLSNMOKDLENSNAKLOEKIQEL 1825
QY 1279 --NEYEVLYLK-----PLAGVYRSLLKQLENNVMTF-----NVNVKDLN----- 1316
Db 1826 KANEHQILITLKDVNETQKKVSEMEQLKQIKQDQSLTSLKLEIENLNAQELHENLEEMK 1885
QY 1317 SRENKRFNFWLESDLIPYKDLTSSVYVVKDPYKFLNKEKRDKFLSSYNIKDSIDTDI 1376
Db 1886 SVMKERNLRRVEET-----LKERDQKES---LOETKARDL 1920
QY 1377 NFANDVLGYKILSEKYKSDLSIKKYINDKOGENEKYLPLFNNIETLYKTVDNKIDLV 1436
Db 1921 EIQOE-LKTARMLSKHEKTEVDKLEKISEKTIO-----ISDIQDLDKSKDELQKKI 1972
QY 1437 IHLEAKVLYNTYKESYEVKIKELNLYLKTITODKLDADPKKN 1476
Db 1973 QELQKKEQLLRVKEDVNMSHKKINEMEQLKKQ---PEPN 2009
RESULT 33
AAB18294
ID AAB18294 standard; Protein; 980 AA.
XX
AC AAB18294;
XX
XX
DT 07-NOV-2000 (first entry)
XX
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:152.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
XX W0200025728-A2;
PN
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
PF
XX 05-NOV-1998; 98US-0107131.
PR
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
XX WPI; 2000-365347/31.
DR
XX
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 362-365; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera of a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rfinos or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the

CC * and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://pub/wipo.int/pub/published_pct_sequences.

QY 1458 KELNYLKTQDKLADFKKNNVGVGIADISTDYNHNNLLTKFLSTGMVFENLAKTVLSNLL 1517
DB 1936 -----NKNQIIG---YSIKYDKN-----VVSNSCSDVITSILK 1966
QY 1518 DGNLOGMLNISQCVKQKQNSGCFRHLDEREC-KCLLANYKQEGDKCVENP 1570
DB 1967 DKKIKR-----KKLQKRN---YENENIVCLDCLLSYLKKMLRIYGNP 2007

RESULT 38
AAB18195
ID AAB18195 standard; Protein; 1516 AA.
XX AAB18195;
AC AAB18195;
XX 07-NOV-2000 (first entry)
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:52.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide.
KW Plasmodium falciparum.
XX Plasmodium falciparum.
OS WO200025728-A2.
XX 11-MAY-2000.
XX 05-NOV-1999; 99WO-US26796.
XX 05-NOV-1998; 98US-0107131.
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 120-124; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are refined or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAB70078 to AAB70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 1516 AA;

Query Match 4.9%; Score 411.5; DB 21; Length 1516;
Best Local Similarity 21.3%; Pred. No. 6.5e-10;

Matches 363; Conservative 262; Mismatches 559; Indels 523; Gaps 92;

QY 141 ELKYP-QLFDTHNMLTLCDNHGFYKL-IDGYEETN-----ELLYKLNFYDLRAKLN 193
DB 37 ELTYANNVKDLNKSF---DNMSIFNDLWDFSENISSEIKTDNIKKAHLYFFFLR---- 89
QY 194 DVCANDYCOI-PF-----NLKTRANELDLVK---KLUVGY--RKPDL 229
DB 90 -ICKLLYNIPIFIFDGNPPPELKRRTIFORNIKRNRVEEKFKAETAKLYVNYOFTLLN 148
QY 230 NIK-----DNVGMEDYIK-----NKKKTEN-----INELIESK-----K 261
DB 149 SMKSNKNDNSNNIEDTNTNKTNTQNSKNTKPNKINADISKSSLIQIYDDIK 208
QY 262 TIDKNKATKEBEKKLYQAQYDLISYKQLEEAHNLISVLE----- 303
DB 209 EKDSLINSILVEHGVNVPVSVKDLVITICNDLSKIKNKIFMITDFGVLFLBGQDGMGV 268
QY 304 ---KRIDLKKNENIKELLDKINEIKNPPPPANGTNPNTLLDKNK--IEHEKEKEIAK 359
DB 269 ENINKLDNRKNDENLSYINRYKQD---VNNNDKDK--DKKENINEVRDOKNYVY 323
QY 360 TIKFNIDSFTDPLEYLYREKN---KNIDISAKVETKESTEPNYPNGV--TV----- 409
DB 324 KNKENLNIIYDDDEKEDIQNKNGVYNNDDIDEQIIRKKHMARKYYEISIKTFKGFIC 383
QY 410 ---PLSYNDINNALNEL-----NSFGDLINPFDTYKTPS-----KNI-YTDN 447
DB 384 MRPVDIIDSINYTEMLEISETLKVHENKFKOHLNVLDENNSTPVVNNMLLNKINYKKN 443
QY 448 -----ERKKEFN-----ETKEKIKIEKKIE-----SDKKS----- 474
DB 444 DDLIEGGEKKSFINLINVDSYSSNSRLNDENTIERGKINFMFITNDEKSIINNYNNNN 503
QY 475 -----EDRSKSLNDITKEYEKLNEIYDSKFN-----NN 503
DB 504 NNNNNNNNDNNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDN 563
QY 504 IDLTNFKMMGKRYSKYKVEKLTHTNTFASY--ENSKHNLEKLTALKVMEYSLRNVIVE 561
DB 564 LDTNFI--FLECKDEYKVVYVVKKEIRIPLFKEINKEIFKLPLOQY---ILQDI--- 615
QY 562 KELYYKNLISKIEN-----EIEFLVENIKKDEQLFEK-KITKDN--APDEKI 608
DB 616 KEETYTDNRKAIKSKDDMDVFSQVQLEYTVRMKTDFE--IEKLKMAENIQSVESGL 673
QY 609 LEVSDIVK---VQVQVLLMNMKIDELKKTQJLLKVELKHNHVPNSVK-----QENK 658
DB 674 LINKDLKNTNDNINIKDYNVLOKKKKKKKKFLNDILNTYNTTESKYQDLYVVGESK 733
QY 659 QBPYYLIVLKKEDIKLVFMPKVESLINE--EKKNIKTGQSDNSEPSTEGEITQATTK 716
DB 734 ED-----IKNQIDFVTOECYRNNDIIRDTHDKSDFIKNIKIDNN-----K 773
QY 717 PQQAGSALGDSVQAQEQKQAPPPVPPVPEAKAQVPTPPAPVNNKT-----ENWSK 771
DB 774 KYEIVNLELEQEBEINEKKNNYNNNDN-----SNKTFELKIENEFK 813
QY 772 LDYLEKLEFLNTSVTICHKYILVSHSTWNEKTLKQYKTKKEESKLSKSDPLDLFLNION 831
DB 814 KDLL-----LDSQIFGDSLLAD-----IKENYTTAD-----NLDN 844
QY 832 NIPVMSYMFDSLNLSLSLFME-----IYKEMVCNLYKLKDN----- 869
DB 845 -----NNENKSLIYEDGENFITRNEPITNEYEKKNNIYISDEQKYNEEDIIFK 892
QY 870 DKIKNLLBEAKKVSIVKTLSSSSMQPISLTPODKPEV-SANDDTSHSTNLSNLSKLFEN 928
DB 893 DKIK---EKEKNNDTSSDDFENCYSQVEKIYVNEKEIEYNNKNDKSSS---SSSILEE- 945
QY 929 ILSLCKNNIYOELIGQKSENFEYKILKSDTFNFENSTFVSKADINSINDSEKSK 988
DB 946 -IKYKKEK--DELV-----SPNLC---VLDEFEHNDLENNTIYSSVDDDKNTVSKNNT 996

Db 525 LPLLCLYICKQNIKAQEBENKTKIKYIILKGGCKQMBENVRRIIHPVLVNLKRCPKYIKIKYLK 584
Qy 514 GKRY---SYKVEK-----LTHHTFAS-----YEN-----SKHLEKLTAK- 545
Db 585 KKHLPISISIKKHLNLTTHDIFLTKQIYVDNYSSQFISLLSLLISFSSKN-----TKL 640
Qy 546 ALKYMEDYSLRNI-----VVEKELK-YKKNLISKIENDEITLV-----ENIKKDEE 590
Db 641 CLNYKHSYTKNMINDVNTYKVIINKQNIYFNNTKNNIKYKIRYLNIHQEKKKKKL 700
Qy 591 QLFEKKITFDENKPEKILEVSDIVKQVQVLLMANK-----IDELKKTOLILKNNVELKINI 647
Db 701 TFFKYMILKKECLLANSILNKLIIPHDCCKGKTMLNQIHLNEENKNDITTKNNNNNNNN 760
Qy 648 HVPNSYKQENKQ-----EPYLVILKKEIDKLKVEPKVESLINEEKKNIKTSG- 696
Db 761 NNNNNNNVNOICVOHKLPCDYTYQNIKKEDYKQGLFNTTSKAFIDMTLYVWRTWGI 820
Qy 697 -----QSDNSEPSTEGEI-----TGQATTTPGOQAG 722
Db 821 HIKVNHKGIVYVQKKEMYQLYDDNNNNNNNNKSDICLNVRNPNKCSSEKKTNP----- 875
Qy 723 SALEGDSVQAQAEQKQAPPPVDPVPAKAQVTPPAVNNKTE-----NVSKLD----- 773
Db 876 -----NSSSILKKDKKKKQ-----MDGKIVTLVKGDNKEEGNNNIKKNDSDASK 923
Qy 774 -YLEKLYEPLNTSYCHRYILVSHSTMKEKILKQYKI-TKEEESKLSCDPLDLFLNION 831
Db 924 GTNEHMQRIND-----AETQNTLHKENKLCITTKQONKIHT-----KINSKE 967
Qy 832 NIPVMSYMFDSLNLSOLF---MEIYKEMVC-----NLYKLD-- 868
Db 968 NEKVKKYVYHINDLGLFYFLVGLFIKKKNTISLKNLNVLNVKYKGNVIKIKTVM 1027
Qy 869 -----NDKIKNLLAEAKVSTSVKTLSSSQPLSLPQDPKPEVSANDDTSHSTNLNN- 921
Db 1028 YOKDIYNTLLNIL-----LLVGVK-----TYIRQHNK-----LNKSEYVNSQNL 1069
Qy 922 -----SLKLF-----ENILSGKKNKIYQELIGOKSENFEYKILKDSDFYN--ESFT 968
Db 1070 IGSKSKSIYVHVITSEISFNKKILRPFYKIQKNNKYKRIIMQSAHINIKESKN 1129
Qy 969 NFVKSKADDINSL-----NDESRKKLEEDINKLKTQLQSLFDLYNNYKUK---L 1015
Db 1130 NIISNNVEKNSVTSNIVSNISSNNISPYKSIKENNMKKTNNCIEHLNNYKIKYNIY 1189
Qy 1016 ERLFDKKTGVGKYKQIK-----KTLLEQLEKSLNLSNPK 1053
Db 1190 EKIIYKTYETNNHMLSFKIVIDAESFDDFSCILFSHFILSNINENIIFKIRNIHQN 1249
Qy 1054 -----HVLQNFVFFNKKKEAEIAETENTLENTKILLKHYGLVKYVINGESSPLKTL 1105
Db 1250 IKESTRIYHVVFILKLFHNL--LFISCTNNSIYITKML--HPLQNIQFYRYKKN-IRTN 1304
Qy 1106 SEESIQT---EDNYASLENFKVLSLEKGLKDNLNLKKSLSYLSGLHHLIAELKEVIK 1162
Db 1305 NOKIYNTYNIHNYEKIQNFVNSKYV-----INDMSLYLYVDTDQNDHRIIFMSTILS 1358
Qy 1163 ---NKN-----YTGNSPSE-----NNTDVN--NALESYK 1187
Db 1359 LIFRNIIPKCDNVHKSPLFFHYAKYKYLHIYVQNSQOFINTYFQDVNNINLHCTKK 1418
Qy 1188 FLPE-----GTDVATWVSESG-----SDTLEOSOPKKPASTHVGAESNTIT 1228
Db 1419 KRPQGSTPDEKYKGEIKGND---LIKESDIKNDIIKESDV-----VNKNEIV 1466
Qy 1229 TSQNVDDVEDDV-----IIVPIGESEEDYDGLGQVYGEAVTPSVIDNIL-SKIENEYEV 1283
Db 1467 ENNNIIEKDEIKTKYTEPI-----KYDN-----TSDAKSISTSTSVLSSESSNELSD 1515
Qy 1284 LYLKPLAGVYRSLLKOLE-----KENMEHNNVITKNNNDNNNNNNNNNNNNNNNNNN 1301
Db 1516 CCMNKL-----KENMEHNNVITKNNNDNNNNNNNNNNNNNNNNNNNNNNNNNN 1569

Qy 1302 -NNVMTFNVNVK-----DILASRFN-----KRENFKNVLESDLIPYKDLTSSNVYVK----- 1347
Db 1570 NNVVEYKPNYKINGLQNIINSCLNFICSKRKNIKNKKIKKIKKKNKIIHNKKKKKNC 1629
Qy 1348 -----DPY--KFLNKEKDKFLS-----SYNYIKDS 1371
Db 1630 NTRHRGNTQINNKLVLINITYILRYPNNNKSSKLSCTKEIKKKTFFRICESYD-IKN 1688
Qy 1372 IDT-DINFANDVLGYKILSEK---YKSDLSIKKYINDKQEN-EKYLPLFLNNIETLYK 1426
Db 1689 IDIHNVNKN-----YKKIDOTLVNHKEEIDTSKQHTDEKICKKIQKY-----LYL 1734
Qy 1427 TVNDR--IDLFFVHL-----BAKVLNNTYKESNVEVKEIKELNYLKT 1465
Db 1735 DVKKRYISLYMYNKKKCKDTNNKIQKKKKEEKKOISYNISSKHSILNRMKY--- 1791
Qy 1466 IODKLADFKNNNFVGIAD-----LSTDYNNH-----NLLTKFLSTGWFENL 1508
Db 1792 --NNIIDMYKRNFTYKDDNYKRIYTYDEILENDINISYLIKQINILNVTICGM--RNV 1847
Qy 1509 AKTVLSNLLDGNLOGMLNISOH-----DGIFCSSNPLGISF 1627
Db 1848 GKTLFUSKKIENNI--IIDIDEYILKDEIKFKLSISDPRIYEVYTFISSLYLAFLITFD 1905
Qy 1531 QCVKQKQCPQNSGCFRHLDEREECKOLLNYKQEGDKCVENPNTCENNGGCDAD-----A 1585
Db 1906 RNLSAPKQDTGATIKHVDIRDE---KINSKQ--NKQTEVDN-DINDNNNNNSDNNHLLH 1960
Qy 1586 KCTEEDSSGKKTKTCECTKPDSPYLF-----KINSKQ--NKQTEVDN-DINDNNNNNSDNNHLLH 1960
Db 1961 NKNQHTSTKKIKQKVSFSDVCEIYVDGPNFENKNYDDNIFITYTN-KGITF 2013
RESULT 40
AAE00425 ID AAE000425 standard; Protein; 2184 AA.
XX AAE00425;
XX 19-JUN-2001 (first entry)
XX P. falciparum telomerase reverse transcriptase full length protein.
DE Telomerase reverse transcriptase; TERT; ever shorter telomere;
KW EST; therapy; stomach cancer; malaria; vaginal candidiasis.
XX Plasmodium falciparum.
OS
XX Key Location/Qualifiers
FH Misc-difference 330 /label= Leu, Ile
FT /note= "Encoded by MTA"
FT Misc-difference 335
FT /label= Asp, Gly
FT /note= "Encoded by GMT"
XX WO200127287-A2.
XX 19-APR-2001.
XX 10-OCT-2000; 2000WO-US27825.
XX 13-OCT-1999; 99US-0417485.
XX (RERE-) RES & DEV INST INC.
XX Long DM, Metz AM, Love R;
XX WPI; 2001-266411/27.
XX N-PSDB; AAD03729.
XX Telomerase reverse transcriptase genes and proteins from Plasmodium

Pr falciparum and Candida albicans are used to detect infections of these
 PT pathogens in mammals -

PS Claim 10; Page 104-111; 138pp; English.

XX The invention relates to identification and use of telomerase
 CC reverse transcriptase (TERT) genes and proteins isolated from
 CC Plasmodium falciparum, Candida albicans and rice. TERT gene is
 CC also called as ever shorter telomere (EST) genes. TERT genes
 CC are used to detect P. falciparum and C. albicans infection in
 CC mammals. They are also used to diagnose the state of an infection
 CC in a patient and the relative amount of the pathogen in a cell,
 CC tissue, organ or organism. The potential binding partners and
 CC modulators of the activity of TERT genes and proteins are used
 CC to inhibit or promote growth of the pathogens and hence are used
 CC to treat e.g. stomach cancer, malaria and vaginal candidiasis in
 CC mammals.

CC The present sequence is Plasmodium falciparum TERT full length
 CC protein.

XX SQ Sequence 2184 AA;

Query Match 4.8%; Score 402.5; DB 22; Length 2184;
 Best Local Similarity 20.4%; Pred. No. 2.5e-09;
 Matches 362; Conservative 268; Mismatches 585; Indels 557; Gaps 94;

Qy 114 SDNSDS--DAKSYADLKHVRNYLLTIKELKYPOLFDTNHLTLCDNIHG-FKYLIDG 170
 Db 148 NDNISDKCITTKNIPKLVHINKYKYLKK-KYHTMY-TNN-----DHSYGKYLIV-- 197
 Qy 171 YEEINELLYKLNFDLLR-----AKLNDVCANDYCYQFPFNKIRANELDLVUK 219
 Db 198 --QCSGRILKNDFFKDMQIOERKYYTSNIIKINSEYTNII--IINNNTNNNNNNNNNN 253
 Qy 220 LVFGYRKPLDNT-----KDNVGMEDYIKKNKT--IENINELIESKTTDK----- 265
 Db 234 NVHGFCH-INNLFPSSNEFPSSISCTWTETKNDKLTTHIRTSLLITENSCKDKLLPEI 312
 Qy 266 -----NK-NATKEEKKKLYQAOYDLISYNKOLEE 294
 Db 313 DFSEDRKESVGYDKKXKNXSNIKRNKINTKEEKKK-----WNKIIN 362
 Qy 295 AHNLLS-----VLEKRI--DTLAKNENIKELDKINEIKNPPANGNPTNLLD 342
 Db 363 RNNILQHTNTNCKTFLFNKHIIFDKIENNIPLFIYDLLNYI-----FKSDQTYFY 414
 Qy 343 KNKTEEHEKEI-KBI-AKTKFNIDSFTDPLEEYV---LREKNKIDISAKVETKE 396
 Db 415 HNNFIDEYKQKICKOIKCTKNDLSHITSRKENHLPVOKLENNYKHPNINKOL---R 471
 Qy 397 STEPNEYPGVTYPLSYNDINNALNELNSFGDLINPFDTYTRKPSNIYTDNERKKFNEI 456
 Db 472 KTKILKY---VINYFKEFINNVIN-TKFGKIYRK-----FPRKHILNKI 512
 Qy 457 KEKIKIEKKKTESKYSYEDRSKSLNDITKEYEKLNLNIYDSKFNNDLNFEMKMGKR 516
 Db 513 HKIFKIIRLQI--IKKYR-----IINRMNRKFIKQVYDT-FFKNYDFLSF----- 556
 Qy 517 YSYKVEKLTHTNTFASYE-----NSKHNEKLTALK-----YMEDYSLRNIVVEKE 563
 Db 557 -SFYTKIINFMVYITRKCIPIKLGSKHNFKIFLKNKYKFLFLPNYKESFSLNQVMKNIK 615
 Qy 564 LK-YKKNLISIKIENIEIETLVENIKKDEQLFEKTKTDENKPDKEKILVSDIVKVOVO-- 620
 Db 616 VNIPOKKISKYINKNRLLKNIFDNN---YENKILHRNK--ELITWINNIKIYNKN 670
 Qy 621 -----KVLMMNKI-----DELKTKQLILKNVELKHNHVPNSYQENKQEPYLIIV 666
 Db 671 DNLNNSFKIKTTLNKLRRKYFNKIKKINIAIQRLANRL---IYLFN---YFTMP 722
 Qy 667 LKKEIDKLVKMPKVESLINEEKNKIKTEGOSDSEPSSTEGEIGQATPKQGOAGSALE 726
 Db 723 LIRRF-----FFLTKSEQTLH---KTIFPDRKIWNH-----FTK-----IS 755

Qy 727 GDSVQAQAEQKQAPVPVPVPEAKAQVPTPPAVNNKNTENVSKLDYLEKLY-----E 780
 Db 756 NFCLYHQIFRNKKL-----KKRNEPKMDIVQNMFNVKKGE 791
 Qy 781 FLNTSYICHKYLIV-----SHSTMNEKILKQYKITKEESKLSLSCDPLDLLFNI--ON 831
 Db 792 KIKT-----NKYIFIKMKKKSTNKINCINFKSKKIPKKKKKN-----LYNITRHN 837
 Qy 832 NIPVWYSM-FDSLNSLSOLFMEIYEKEMVCLVKLKDNDK-----IKNLEBAKYS 883
 Db 838 NIFIKKMEKSKTNNLIN-----KSDNLTKLEINKKSVRPYIKKYYIKKKY 888
 Qy 884 TSVKTL-----SSSMQPLSLTPQDKPEV-----SANDDTSHSTNLN 920
 Db 889 FALKKMYIHRMAKEEKNIKLERAFKHFIFAQEKELHILKYFSSHFFQNRKINYGKRN 948
 Qy 921 NSLKLFENIL-----SLGKNKN-IYQELIGOKSSNFYEKILKDSOTFYNESFTNFVSK 974
 Db 949 KLIHRIKNIITIKQNSGIVKNKDKTFLHLIKNKSNNNNK-KKNKNYNNNNIN--NNN 1005
 Qy 975 ADDINSLND-----ESKRKLEEDIN--KLAKTLQLSPDLNLYKYLKLERLFDKK 1023
 Db 1006 NNNNNNNNNNNCKLSNKRNYRNNNNNNKAKNKNNDSDN-----LEK--KK 1057
 Qy 1024 TVGKYKMQIKKLTLLKEQLESKLSLNNPKHVQLQNFVFNKKKKEAEIAETENTLENTKI 1083
 Db 1058 KIYIVKI--KNIIEKRNFMKLNSINH-----FISKLRIN----- 1091
 Qy 1084 LLKHYKGLVYNGESSPLKTLSESIOTEDNYASLENPK-----VLSKLEGLKD- 1134
 Db 1092 WIPKKKGLRPLIN-----LSTLNVPFIVKQRFITELKSKSEFYFHNILNLEREKDK 1146
 Qy 1135 NLNLEKK--KLSYLSLSSLHLHIAELKEVKNKNYTGNSPENNTDVNNALE--SYKKFLPE 1191
 Db 1147 NIKRRKKNKNFNPVSLNNTCNFSLKCLGNRRHNNNSLFKNTLTKTGEIELKLLKWL-- 1204
 Qy 1192 GTDVATVYSESGDTLEQSPKPKPASTHVGAESNTITTSQNVDDDEVDDVIVPIPESEBE 1251
 Db 1205 -----HYLKNWFYKKRKKKIYI--KNLKNKKKIYAI-----CIGDSN 1242
 Qy 1252 DYDDLQGVVTEGAVTPSVIDNLSKIENEYEVLYLKLPLAGYVRSLLKQLENNVMT-FNVN 1310
 Db 1243 CYEHINHNLYPK-ILKNFEDNI-----NNFEIYL--FKRSPRLYNKLNLSFLSYYPVN 1294
 Qy 1311 VKD-----ILNSRENKREN-----FKNVLESDLIPYKDLTSSNVVWDPYKF 1352
 Db 1295 VKSGLHYIRNLRELIIKSHLNDNHHLNOMFKTKSKDL-----YIFADSYKS 1344
 Qy 1353 LNKEKRDKFLS-----SYNYI-----KDSIDTDINFANDVLGYKILSEKY 1394
 Db 1345 LOVDKRDIFMTIITVIRYIYLNIIYFSIKFEKLNRKNIFVFOENOMKGVY--LSVRDK 1402
 Qy 1395 SLDLSIKK-YIND-KOGENEKYLPPLNNIETLYKTVNDKIDLFTVHLAKVLNLYTEKSN 1452
 Db 1403 KRVENIKWYLNMSKKINHDEITLESKNSI--NINKNFMICTNHEOD-----TEEGN 1455
 Qy 1453 VEVKIKELNYLKTIO---DKLADFKKNNFVFG-IADISTDYNNHNLTKFLSTGMVFEN 1507
 Db 1456 TONKEKHDIYIGPIYNNNSFDTSTTTTHSSNNYKGNINHVGSDYKNDGLLHGK----- 1506
 Qy 1508 LAKTVLSLLDGNLQGLMLNIHQHOCVKQCPQNSGCFRHLDERECCKLLNAYKQEGDKCV 1567
 Db 1507 -----NNSMNECYVKDI-----KCN 1521
 Qy 1568 ENPNPTCHENNGCDA---AKCTEEDSGSN---GKKTCTCTKPDSPYPLPDGIFCSCS 1620
 Db 1522 NNNNNNNNNNNNNNSYNKLCNVTNNSKNDIYKHTIDTDSKNHTY--FKNKFL--- 1576
 Qy 1621 NPL-----GISFLLILMLIYSFI 1639
 Db 1577 NFLOKRIISNIYGLPQGFSLNLSILGSLYAYL 1608

RESULT 41
AAB18265
ID AAB18265 standard; Protein; 2013 AA.
XX AAB18265;
AC AAB18265;
XX 07-NOV-2000 (first entry)
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:122.
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoa; infection; insecticide.
XX Plasmodium falciparum.
OS
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 285-291; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 2013 AA;
Query Match 4.7%; Score 399; DB 21; Length 2013;
Best Local Similarity 19.5%; Pred. No. 3.2e-09;
Matches 366; Conservative 235; Mismatches 593; Indels 684; Gaps 85;
XX 157 LCDNTHGRKYLIDGVEEINELL-----YKLNFFDLRLAK----- 191
DB 15 ISDNI--FCFLKDGVCIFCNLLNNEKKYLYITCSQDEGVYQYVDFVVKRYEKREDCN 72
XX 192 -----LNDVCAN-----DYCQIPFNKLIRNELDV 216

Db 73 KNMTINIMLLQNEKKIIKETCYIKNVVTNKTYYHTFLVINKHYHNILCSLSFENNPF 132
---Qy 217 LKLVFGYRKPLDNIKDNVGMEDYIKKNKTPNIENINELIESKKTIDKNKATREEEKK 276
Db 133 LN-----TNFVKTFGK-----IKSWACTNNNIFVLIIKKKKLIINKN---NQMS 176
Qy 277 KLYOAYDLSYINKQ---LEBAHNLISVL-----EKRIDTLKNE-----NIKEL 318
Db 177 KILNQNLVSKHTLDRSLLLMKGENDVNVICBSKKEKKKKKKTKDNKNEKKKGHMEIKDV 236
Qy 319 LDKINEIKPPANGNTPTLLDKNKKIEEHEKEIKETAKTIKFNIDSLFTDPLELEY 378
Db 237 NEKINEKINE-----EKNEKINEEKNEKNEKIN-----EKK 272
Qy 379 LREKKNIDISAKVETK---ESTEPNEYPNGVTPLSYNDINNALNELNSFGDLINPFD 434
Db 273 NEEKNEKINEEKNEEDTKDPEEKENDNIPGDHSVQYNIPTFSI--LNK----- 321
Qy 435 YTKEPS-KNIYTDN---ERKKFINEIKIEKKIESDKSKSYEDSKSLNDITKEYE 489
Db 322 --KEPDLKKIQFSNIILPIKKMIICPYDEKIIII---LLSHKSIIVITITKNNDLKNMF 375
Qy 490 KLLNEIYDSK-----FNNNIDLTNPEKMMGKRYSYK-VEKLTHTHTF-- 530
Db 376 IIKELIFNSPIITTTWIDNYIFLIYFLNNELIFLSPAKPCRNLYFYKCNINNYSHITSF 435
Qy 531 -----ASYENSKHNLE 541
Db 436 KSRNLYISFKTEIVCFKIRYVEIPLTVPKKIQTTEGVIDAKYLFERKPRVINTINHNOS 495
Qy 542 KL-----TKALKYMEDYSLRN-----IVVEKELK-----YKKNLSKTIENIETLVENI 585
Db 496 NAKDKDGDNDVIREEDF--LRNNKNFSDVKRRKRNKNKNNYEIIFNNILREIKTLENKI 554
Qy 586 KDEQLF--EKKITDKENKPKDEKILEVSDIVKQVQVLLMKIDEKTKTOLILKNVEL 643
Db 555 SNNDYNIETDEGEINKDELKN-----RLSARSLSVYKNTFNLML-----L 594
Qy 644 KNIHVPNSYKOEKOEPPYL-----IVLKEIDKLKVPMPKVESLINEKNI----- 692
Db 595 GHNNKKKWIRODIRNMVHNKYNCVEEDVCINRYEKESIFV-----EYDNNNDNMLWS 650
Qy 693 -----KTEGSDNSEPSTEGETGQATTKPCQQAAGSALGDSVQAQAOEQKQAPPVP 745
Db 651 HLYFLKKKKKKFFEDHYNDENVI-----KLLDFVSI----- 682
Qy 746 VPVPEAKQAPPPAPVANKTENVSKLDYL-----EKLYEFLNLS--Y 786
Db 683 -----INLHKYILANNITSFYIMSKYLFVLLDGLLYTTKKNDGKIYDFLELSNFY 733
Qy 787 ICH-----KYVLVSHSTMN-----EKILQYKITKEE 814
Db 734 ICYKYNKINKVIDIKINEHDIVYMDKKHILKNHSLKNYLNINTIKEI-QSYNIF---- 788
Qy 815 SKLSCDPLDL-----LFNIQNNIPWYMSFSLNLSLSQLPWEIYKEKEMVCNLYKL 866
Db 789 SMLENTCITFLSLNDGSPFYINITKHKILLY---ENLQN-----FSNLGHNIQYCNFKN 840
Qy 867 KD-----NDKIKNLEEAQVSTSVKTLSSSMOPLSLTPQDKPEVSAANDDTSHST 917
Db 841 KYIQSVFNKLNNEYIFNGFYVQOYIIFFFLIYSTSHKFFIYLVENITHIYLFKRIHOT 900
Qy 918 NLNLSKLFENITSLGKNKNIYQELIG---QKSSRNF-----YEKILKD-----SST 961
Db 901 NI-----LVKN---KEKNQONENIINMKROKESNYILYNYLYKTLNKNKYVCLCSDK 952
Qy 962 ----FY-----NESTNFV--KSKADDINSND-----ESKRKK 989
Db 953 SVSYFYMFEDLPREEEIKMYISDKKKKKKINNNDNKKYIYNRKNKONNNYKENQKNE 1012
Qy 990 LF-----EDLNK-----LKKT-----LQLSF-----DL 1007
Db 1013 VENHYDDDDDDNKSPLYTRNIFFCISKNNTNIVYAKICIGNTWIVADYVLTNITFYIKDN 1072

SQ • Sequence 2295 AA;		Query Match 4.6%; Score 388.5; DB 21; Length 2295;		Best Local Similarity 19.9%; Pred. No. 1.1e-08;		Matches 370; Conservative 282; Mismatches 591; Indels 615; Gaps 98;	
Qy	178	LYKLNFPD--LLRAKLDVCAQYCIQIPFLKIRANELDVLKLVFGYRKPLDNIDNV	235				
Dy	5	VFLNFIFFPFLLIKDESVNGRKEIYF-----DDEKLKLSFPDRS--TNINLDV	56				
Qy	236	G---KMEYI-----KKKKTIENINELIESKKTIDK--NKNATKEEKKLQAO	282				
Dy	57	GENDESSYVPREVDEKKKKKKIDS-----KENSNGNIYKNDTKNNE-----	106				
Qy	283	YDLSTYNKOLEAHNLISVLEKRDITLKKNEINIKELDKINEIKNPAPPANGTPTLLD	342				
Dy	107	YNVVLKGRAREG---IITDEKRSSTKGRNKEQNNKNSDDVDHNNMNDINFVVE	163				
Qy	343	KNKIEEHEKEIKE--IAKTIFNIDSFTDPLELEYILREK--NKNIDISAKVETKE	400				
Dy	164	YKMDYNDYKILDELILKSIINRYNYFN--MLDEYSLOTKLKEM-----	207				
Qy	401	NEYNGVTYPLSYDINNALNELASFGDLINPFDYTRKPSKNI--YTDNERKKFI--	457				
Dy	208	-----YDSLNYLILMNN-----KNSRYFISFSNNEKKKIKNDMN	244				
Qy	458	EKIKIEK-----KKIES--DKKS---YEDRSKSLNDITKEYEKLILNEIYDSKF	500				
Dy	245	ENIYIRHFIVSLFRWYNFKLIECFCKNNFIYIDENK--IYSKYNYKMLNLFSSNF	303				
Qy	501	NNIDLTNFERK---MGKRYSYRKEKLT-----HNTFASYN---SK	537				
Dy	304	LYIYNLSKSLLEIIDNYKYSFIINNIRKDYPNMYVCQSYFDYFYSYFLSYNHHF	363				
Qy	538	HNLEKTKALYMEDYSLRNIVKE-----LYKYNLISKIEIEIETLVENIKKDE	592				
Dy	364	H-----KYLINMDIWNNSIQTKQIGNHKLYKLLKLNENLI--LYNYIKND	412				
Qy	593	FEKKITKDENKPEKILEVSDIVQVQKVLNKNKIDELKTKTLKVELKHNHVPNS	652				
Dy	413	-----IPVTLERMI FSNF-----TNLLIDILNKLYNIDYQDN	446				
Qy	653	YKQEN-----KQPPYLIIVLKKEID--KLK--VFMP--KVESLINEEKNKIK	699				
Dy	447	IKQENVNVPQDAPQDVYHNKNDVSLKNVKEPKVEH--NKAMSNEYTERDGD	504				
Qy	700	--NSEPSTEGET-----GOATTPGQQAAGSALEGSVQAQAEQKQAO	741				
Dy	505	NTNKEFEKSEGTNNISGGEDTFKNISGGEDTFKNISEGDEYDGDG--EGD	563				
Qy	742	PPVPVPPEAKAQVTPPPAPVNN-----KTENVSKLD--VLEKLYEFLNT	789				
Dy	564	DDSSVDTHNNKNDGKESDVNLLMSYKKLANDENFKYKYLKMLDRLKNSSEKK	623				
Qy	790	KYIIVSHSTMHEKILQYKTIKTEESKLSSCDPLDLAFNIQNNIPVMYSMFD	847				
Dy	624	EDI---NSYKNYELKEGIIYNYKVSQY-----IPLIFNPTKDV-----	669				
Qy	848	SQI--PMEIYKEMVNCVLYKLDNDKIKNLLLEAKVSTSVKTLSSSSMOPLSL	906				
Dy	670	SKINFFNIY--EYIITITPKYEN---KNFYODLLKCRREIFKDRHLLENN	722				
Qy	907	VSANDTSHSTNLNLSLKFENILSLGKNKI-----YQELIGOKSSENFY--	956				
Dy	723	-----ELKKNIRNLMRIHE--VSNEGNRRNTINRKYYKYTYDYDKMNEL	774				
Qy	957	--KSDTFYNESFTNPFVSKADDINSLNDESKRKKLEEDINKLKTQLSOLF	1015				
Dy	775	VNDTNTF-----NFMNKEKDKNYF-----DIN---KTMRI--DY	813				
Qy	1016	-----ERLFPKKTIVGK---YKQIKKLTLLKEQLESKLSLNNPKHV	1061				
Dy	814	FTPAAIKMKDIYDQKILLRSNFVEKLKNSICVLSFLYLIGINDDGKLF	873				
Qy	1062	FFNKKKEAIEAETENTLTKILLKHVKLVKYNNGESSPLKT--LSEESIOTED	1115				
Dy	874	DFSVKL---IREKDGKLCNFLSGVLYHINLPIFVNNSSISISTEMDDVLE	930				
Qy	1116	-----YASLENF-----KVLS-----	1126				
Dy	931	FYIYKNNENIRNHDLSDENRIIPRKEDNIKSIISYSLGSSKDDFFSKLA	990				
Qy	1127	KLECK---LKDNLNLEKKLSY-----LSSGLHLIAELKE---	1159				
Dy	991	KYKNTNTNYLKDYDFDFTDKINKSVIKNNVSPFLTCDYLLSNILGAVVDS	1050				
Qy	1160	-----VIKNKN-----VTGN-----	1169				
Dy	1051	LESGVYEENINDKKNIIQNTVQVQKNLFEYFVKLADNRNSYALAALEI	1110				
Qy	1170	-----SPSENNTDVNNA-----LESYKKFLP-----EGDVATVVS	1203				
Dy	1111	ERDEIKAFEFWKKAADQDGTTSALSTGYAYLDEYKFKLKEELVKNM	1170				
Qy	1204	SD-----TLEOSQPKRPASTHVGAESNTIITSONVD--DEVDDVIIV	1256				
Dy	1171	STOKKNVTLEMFQESSEKKNQKKKEKQDQDGTGDRVDDKIVQVNV	1230				
Qy	1257	GO-----VVTGEAVTPS--VIDNLT-----SKIENEYEV---LYLK	1302				
Dy	1231	DESMGRNGSIDCFMPPSGGLNNVSVQNNANIONNANTONNANIQSN	1290				
Qy	1303	NVMT-----FNVNKDILNSR--FNKRFNKNV--LESDLI-----	1342				
Dy	1291	NIQSNANIQSNANIQSNVNSHGTTNRQNNINNVNFENAVTQOTS	1350				
Qy	1343	NYVVKDPKFKLKEKRDKFLSSYVIKDSI--DTDINFAND-----	1381				
Dy	1351	SFSSVSPSSFLFDIPEG---SEYEHTENILDEQNMFTNKNKEQEGGPN	1407				
Qy	1382	-----VLGYKYKILSEKYKSDLSIKYI-----NDKQGEN--EKY	1421				
Dy	1408	DENDEMIKKYKMDLNDLNLKSLKNAEYFHKAIRNDDSDLENILAKY	1467				
Qy	1422	ET--LY--KTVNDKIDLFVHLEAKVNLNTYEKSNVEVI-----	1462				
Dy	1468	ELAGIYKKAADKGD-----NISQMLLGHYSGSDIGIKLNDYKDD	1523				
Qy	1463	-----LKTQDKLADFKNNNFVGIADLSTDYNNH--NLATK	1497				
Dy	1524	SAQNGNIISLYNKSILILKGVNPKYKTFNKECKTKLHFHFTGL-----	1577				
Qy	1498	FLSTGMVPEP-----LAKTVLSNLIDGLNQLMNLISQHCVKRKCQ	1540				
Dy	1578	LLRRNYQFKDYTGSLLSLIMSEL--GDHAHNVNASMLWTLLKRTMQ	1635				
Qy	1541	SGCFRHLDEREECKLLNFKQEGDKCVENPNTCENNGGCDADAKCTE	1598				
Dy	1636	LSLIKELKNKEE-----KEKEKKNKNIHNVNNNSNINGYKKC--	1684				
RESULT 44							
ID	AAV31946	standard; Protein; 4134 AA.					
XX	AAV31946;						
XX	21-DEC-1999	(first entry)					
XX	Plasmodium falciparum	cytoadherence gene protein GLAG9 paralogue.					
XX	CLAG9; paralogue; cytoadherence linked asexual gene; CLAG;						
XX	erythrocyte; red blood cell; malaria; infection; therapy; vaccine.						
XX	Plasmodium falciparum.						

PN W09949048-A1.
XX 30-SEP-1999.
XX 25-MAR-1999; 99WO-AU00213.
XX 25-MAR-1999; 98AU-0002580.
XX (MENZ-) MENZIES SCHOOL HEALTH RES.
XX Kemp DJ, Trenholme KR, Gardiner DL, Holt DC, Cowman AF;
PI WPI; 1999-591099/50.
XX New proteins useful for treatment of Plasmodium infections in humans,
XX especially malaria
XX
XX Disclosure; Page 90-115; 150pp; English.
XX
XX This sequence represents the protein product of a clag9 paralogue
XX gene (see also AAZ20056) on chromosome 3 of Plasmodium falciparum.
XX It is a paralogue of the newly identified cytoadherence linked
XX asexual gene 9 (CLAG9) protein (see AAY31945) encoded by the clag9
XX gene on chromosome 9 of P. falciparum. CLAG9 facilitates
XX cytoadherence of erythrocytes parasitised with Plasmodium to other
XX cells. The identification of clag genetic species, and the
XX products encoded by them, enables a range of therapeutic agents to
XX be rationally designed and/or identified that are useful for the
XX prophylaxis and treatment of disease conditions caused or
XX exacerbated by infection with Plasmodium spp., e.g. malaria,
XX especially in humans.
XX
XX Sequence 4134 A4;
XX
XX Query Match 4.6%; Score 387.5; DB 20; Length 4134;
XX Best Local Similarity 21.1%; Pred. No. 2.4e-08;
XX Matches 357; Conservative 245; Mismatches 525; Indels 567; Gaps 93;
XX
XX 134 NYLLTITK-----ELKYPOLFDTNHLTLCDNIHGFYLDIGYEINELLYKLFNY 184
XX 114 NYLNTVKVNSFSSWVELKEQORERLFHYNMT-----KGVVSFIDSFEHKGPNTHICMV 168
XX
XX 185 FLLLRALKNDVCAN--DYCOIPNL--KIRANEL-----DVLKLV 221
XX 169 FEFMGPNLLSLKHVDYKGIPLNVRKIATHVLIGMOYLHDVCKIITHSDIKPENLVLSPL 228
XX
XX 222 FGVRKPLDNIK-----NVCKMEDYTKNKKTI-----ENINELIESKKTIDKNKA 269
XX 229 TTIPRPRDTYTKDLESKNQVEKKENDQNVCKLITTTNNNTNINTNLSEKKVI-----ND 284
XX
XX 270 TKREEKKLYQAQYDLISYKQLEEAHNLISVLEKRIIDLTKKNEIKELDKKINETKNPP 329
XX 285 TOKNDR-----NIEYDQKTSK-----ENIEDNVSVFNDPSDPN 319
XX
XX 330 PANSGN---TPNTLLDKNKKIEHE--KEIKETAKTIKFI-----DSLPT 370
XX 320 QKNLNNNTIDNNIIPSNVQIEKQSTLSKNKKNEKSDSYININSLTNDQNLKREDIKEN 379
XX
XX 371 DPLE--LEYLREKKNIDISAKVETKESTEPNGVYPLSYNDINNALNELNSFGD 428
XX 380 DXAEGITKYDMLNLIKNNISIKERINDCHS--PENKN-----KDNHQCED----- 423
XX
XX 429 LNPFDYTPKPSNIYDNERKFFNEIKETI-KTEKKIESDKKSYEDRSKSLNDITKE 487
XX 424 --NSINICNKNNTIOTNIND---NTVNEKINNTSKKDLNNTQNNDSEK--NDVIVIE 476
XX
XX 488 YEKLNEIYDSFNNNIDITNEKMMGKRYKYKVEKLTHNT-----FASYENSKNLEKL 543
XX 477 -QQLVNEEDILKKKNQK---TKKKKNINEP--PYVKHKLSPNSDPSLLTSYSNIHALQETL 531
XX
XX 544 TRALKYMEDYSLRNIVVEKELKYKKNLISKIENIEITLVENIKKQEBQLEFKKTIKDNK 603
XX 532 TRKPYHNTYFLNN---PEK-----YR-----DNKMNPYLHRLPND-----CLKKIDQDDSD 575

QY 604 PDEKILEYSDIVQVQVLLMKNKIDELKKTQILKLNVELKHNIHVPNSYKQENKQEPY 663
DB 576 ETEEDDLSDVDQ-----NK-EQKNQL-----EVNLPNN-KYPNSNDVYK 614
QY 664 LIVLKEIDKLKVPMPKVESLINEE-----KNKITE-----GQSDNEP--- 703
DB 615 F--FEKDINKFPIYCDMFNHLIHPALRHELKLYEFLNTSYICHKYILVSHSTM-NEKILKQ 806
QY 704 -STEGEITCOATTKPGQQA-----GSALEGD-SVOAQAOBQKQAQPPVPV 747
DB 673 INTE--DGEYCIPIYDPFVSVYHESKCYKICDLGNSLWIDESRYAEIQRQTRAEVIL- 728
QY 748 VPEAKAQVPTPPAPYNNKNTENVSKLDYLEKLYEFLNTSYICHKYILVSHSTM-NEKILKQ 806
DB 729 ----KSGF-----NETADI-----WSFACMVFEVLVTGDFLFPQ--KG 760
QY 807 YKITEESKSSCDPLDLLFNQNNIPVMYSFMSLNSLSQLFMEIYEKEMVCNLYKL 866
DB 761 DRYDKNEEH-----LSFIEVLGNIP--KHMDAGYNS-----HKYFNKNMYRL 802
QY 867 KDNKIK-----NLLEAKKVSTSVKTLSSSMQPLSLTPQDKPEV----- 907
DB 803 KNIRIKKYGLYKILKYKNLPEK-----EISPLCSFLLPMLSDVPQTRPSAYTMLQHP 856
QY 908 -----SANDDTSHSTNLNSLKLFLNLISLGLKNKNIYOE----- 941
DB 857 WLMNYSLEEGDDMYINDESYISNDRNMKNNSNNFIYDGHNS-SKNKSNKKKIDVN 915
QY 942 -LIGQKS-----SENFYEKILKSDTFYNESFTNFVKSADD-----INSLND----- 983
DB 916 YKIGNGNAYNDNYNKNYKNKN--NKNFNDVDVPEPSDQMYHANYNDLVHVLVYK 973
QY 984 -----ESKRKLEEDINKL-----KKTQLSFDLYNKYKLLERLFDKXK 1023
DB 974 PYSNNVISTYNNKGHKNFDLYLOHRDNNSNQNTSLTNDTFNSDYLTANMDH-- 1031
QY 1024 TVGKYKMQIKK-----LTLKEQLESKL--NSLNKPHVLQ-----NFSVPFNKKKE-- 1068
DB 1032 --DTRYKQIIKIIPAQISKLKDGKNFAYNESIQVHDFQYNEHDFEYKFNKRFHA 1089
QY 1069 -----AETAENTLE--- 1079
DB 1090 HHKEMKHNDDEYEBEDEDEDDDEDYEDVDYDDDEYDEGOEHADADOKNNNEKQ 1149
QY 1080 -----NTKLLKHVGLKYNGESSPLKTL--EESIQOTEDNYASLENEFKVLSK----- 1127
DB 1150 EQONTGEKYNIEHNNMGY-----NKNIOQLSYTNNDNENFCETQNIYIILQNRDINF 1205
QY 1128 LEGKLKDLNLKELKKLSYLSGSL-----HHLIAE-----LKEVIKNKNYTGNS--PSN 1174
DB 1206 KECTPRNNINKEIKSKDYQSSKVINQKONYWYKIKENTKLREHAKKHQYSSNNNNINKND 1265
QY 1175 NTDVNNALESYKFFPEGTDAVATVYSESGSDTLEQSOPKKPASTHVGAESNTITTSQND 1234
DB 1266 NTNIMNQIDTQDQISKNLHDLST-----NNNDQK-----HGAJQKMHMNEKTND 1311
QY 1235 DWDVDVIVIPFGESEEDYDGLGVVTGEA-----VTPSVI-----DNIL 1274
DB 1312 KPLNDEEL-IENRDQNVNKNCKVINKNSCAYTKWVSFFKTIPIIFFLLCLNEKVL 1370
QY 1275 SKIENEYVLYLKLPLAGVYRSLLKQLENVMTFNVNVKDIILNSRPNKRN--FKNVLES 1331
DB 1371 CSI-NENENLG-----ENKNENANVTNENLKNLLNEYDNIEQLKSMIGN 1414
QY 1332 DLIPKDLTSSNVV-----KDPYKF-LNKEKDFE--LSSYN-----YIKDSI 1372
DB 1415 DEL-HKNLTILBKLESLEKDKLKYPLKQGTQELIDISKFNKNKNTADDEETIIPV 1473
QY 1373 DTDINFANDVLGYKITLSEK-----YKSOL-DSIKKYINDKQENKYLPLPFLNIETLYKT 1427
DB 1474 QSSF---HDIVYEHLEKQSEIYNSDLSKIKKI-----FI--VRTL-KT 1515


```
Db 669 G E F F F D E V E E K F D E K M G E F F D E V E E K F D E K M G E N I F E I P K -----K D V E I E E T Y S E K 723
Qy 1015 L E R L F D K K T V G K Y K M Q I K K L T L L K E O L E S K L N S L N N K P H V L O N F S V F F N K K K E A E I A E T 1074
Db 724 M W E I H E K -----I K D K D E P E K I H D E K K E V E F F L I A D R K K E -----762
Qy 1075 E N T L E N T K I L L K H Y K G L V K Y Y N G E S -----S P L K T L S E E S I Q T E D N Y A S L E N F K V L S K L E G 1130
Db 763 E N E D S N V E I - L N I D K N N F Y F E N K E T F E I D E K V S K M E E D F V Y E N N -----E T F E - C E D I F L 816
Qy 1131 K L K D N L N L E K K L S Y L S G L L H L I A E L K E V I K N K N Y T G N S P S E N T D V N N A L E S Y K - R F L 1189
Db 817 K R E D N D S E N E K -----E I D E I G E V I N I G I Y H L N K N S Y D D V H I L T H D F K N E L L 866
Qy 1190 P E G T D V A T V W S G S G T L E Q S O P K P A S T H V G A E S N T I T T S O N V - - - D D E V D V I I V P I F G 1247
Db 867 I E K Y N V D I C S D -----D N I Y D G N I C D D N I Y D -----G 896
Qy 1248 E S E E Y D D L - - - G Q V V T G E A V T P S V I D N I L S K I E N E Y E V L Y L K P L A G V Y R S L K K O L E N N V 1304
Db 897 D N I Y S G D N I Y G D N I Y S G -----D N I Y S G - D N I Y S G D N I H S G D N I Y S G -----D N I 941
Qy 1305 M T F N V N V D I L N S R E N F K N V L E S D L I P Y K D L T S S N Y V V K O P Y K F L N K E K R D K F L S S 1364
Db 942 D D N I Y D G N I N S - - - G D V E N L K E H K I A V N E S E E T A Q D I K E Y - - - E K R D N E F T - 991
Qy 1365 Y N Y I K D S I D T I N F - - - - - A N D V L G Y K I L S E K Y K S D - - - - - L D S I R K Y I N D 1406
Db 992 - D Y V E E - - N S D I R F Y K G K G E M V N E L I G E Y - - - S E K Y M D N N I E D N E L V I W S A S V K - - - N D 1042
Qy 1407 K O G E N E K Y L P F L N N T I E T L Y - K T V N D K I - - - - - D L F V I H L E A K V - - L N Y T Y E K S - - - - - N 1452
Db 1043 K E R L A D D N I D L N N N I S N D Y I K N N E D I K N V H D S F S I S K S E L H D I N G L E I K S I S S N D I K S 1102
Qy 1453 V E V K I K E L N Y L - - K T I O K L A D E F K K N N F - - - V G T A D L S T D Y N H N L L T K F L S T G M V R E N 1507
Db 1103 I E V C V K E N E I H H K M M K K E L N D N L N D E M Y W C D I S N D I F K N N E Y T K H V D D V Y T D - 1161
Qy 1508 L A K T V L S N L L D G - - - - - N L O G M L N I S O H Q C V K K O C P Q N S G C F R H L D E R E C K C L I N 1558
Db 1162 - - E N N S N L I I G E D E H C V S S M N F E Y P F N I S - - - - - K M T E S N N I L Y E Q N D K - - - - - 1205
Qy 1559 Y K Q E G D K C V E N P N P C N E N N G C D A - - - - - D A K C T E - - - - - 1589
Db 1206 - K K T I N S V K H P M Y I K G F A S D S I N F E L K A L K G L P P L P F L K C K D M K P Y M R F N I V L K V I 1264
Qy 1590 E D S G S G K K I T C E C T K P O S Y P L F D G I F C S S N F L G I S F L L I L M L I L Y S F I 1639
Db 1265 E S N D Y N G R R - K I K V T K M - - - - - F I C L K L K F D M I Y V I I Y I L Y I F L 1305
```

RESULT 47

AAB18234

ID AAB18234 standard; Protein: 2573 AA.

XX

AC AAB18234;

XX

XX 07-NOV-2000 (first entry)

XX

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:91.

XX

XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;

KW

XX antimalarial; malaria; protozoacide; infection; insecticide.

XX

OS Plasmodium falciparum.

XX

XX WO2000025728-A2.

XX

XX 11-MAY-2000.

XX

XX 05-NOV-1999; 99WO-US26796.

XX

XX 05-NOV-1998; 98US-0107131.

XX

PA (HOFF/) HOFFMAN S.

PA

PA (CARU/) CARUCCI D.

PA

PA (GARD/) GARDNER M.

PA

PA (VENT/) VENTER J C.

XX

XX Hoffman S, Carucci D, Gardner M, Venter JC;

PI

XX WPI; 2000-365347/31.

XX

XX Disclosure: Page 213-220; 577pp; English.

XX

CC The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum. Also described are: (I) nucleotide sequences (II) encoding (I); and (2) vaccines against P. falciparum infection comprising (I) or (II).

CC

CC (I) and (II) are useful for the development of vaccines against P. falciparum infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with P. falciparum. Furthermore,

CC

CC (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent P. falciparum infection, or they can be used to identify drug resistance in P. falciparum. Sequencing of the plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand

CC

CC our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

XX

XX SQ Sequence 2573 AA;

Query Match 4.6%; Score 383.5; DB 21; Length 2573;

Best Local Similarity 18.7%; Pred. No. 2.1e-08;

Matches 359; Conservative 288; Mismatches 625; Indels 643; Gaps 86;

Qy 119 DSDAKSYADLKHVRNLLYLLTIKELKYPQLDLTNHMLTCLDNIHGFKYLDIGVEEINELL 178

Db 189 ENKKDEKEKKNIHNN -----DENNNMI-YKKNIDKTHYILD--NNVVHL 232

Qy 179 YKLNPFY----DLLRAKLNDVCANDYQIPFNKIRANELDVLKLVGYRKLPLDNKDN 234

Db 233 NDINTYLRERDYMNRKFGTYIDSTY-KNPMYVTVLYIFNNDILKDIL---QVIDIIRD 288

Qy 235 ----VGKMEDYIKKKTIENINEL-IEESKKTID--KNKNATKEEKK-----KLYQAO 282

Db 289 FDHAIYKDIDENOLIKNLIILNHLTTRPSKEWFDYKWRHMTFNDKKSEYNYKYLOQ 348

Qy 283 -----YDL-----SIYNKOLEEAHNLISVLEKRIOTLKKENIKELLKDKINEIKNPPAN 332

Db 349 KSDRILYDLTKNDIYIKELQKRSIDLOYQKGLQSLKCLLANKNFLTMLNFERY----- 403

Qy 333 SGNTPTLL-----DKNKKIEE---HEKEIKETIKFKNIDSLTFDPLELEYLR 380

Db 404 -----NTQLFDADYRETEENKVMQORRELELLEKKRLK-----QELESYHD 448

Qy 381 EKNKNIDISAKVETKESTEPNPGVYPLSYND-INNALNELNSFGDLINPFDTYKTP 439

Db 449 DSSTDDSSADQDQDERE-----VLTHNDPINKKDDPIN----- 490

Qy 440 SKNIYTDNERKKFINKIEKKIKIEKKIESDKSYEDRSKSLNDITKEYEKLLNEYDSK 499

Db 491 -KNDDPINKNDNDINKNDNI-----NKDNDICNSDHTNSNDHTNSNDHTNSNDHT 545

Qy 500 FNNIDLTNFEK---MMGKRSYKVEKLTTH-----HNTFASYENSKH-----NLEKLT 545

(1) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAT70078 to AAT70287 and AAT8144 to AAT8352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

AA	Sequence	2206 AA:
SQ		

Query Match 4.5%; Score 379.5; DB 21; Length 2206;
Best Local Similarity 19.9%; Pred. No. 2.6e-08;
Matches 294; Conservative 214; Mismatches 451; Indels 515; Gaps 72;

QY 240 DYIKNNKTTIENINEL----IEESKKTIDKNKNATKEEEKKKLYOAOYDLSIYNKOLEEA 295

Db 2 EYLESEKSSDDRREVNNFENDYSKDSSSHNTNSDLVDVRKKHSDNVYESEQDGKQTEG 61

QY 296 HNLISVLEKRIDTLKKNENIKEILDKINEIKNPPPPANSNGTPTNLTLDKNKKIEEHEKEIK 355

```
db      |          ||| ::  
Db      |62 RKKI---KGFFLKKGDS-----EDENKEKETKDHLRK 91
```

QY 356 EIAKTIKENIDSLFTDPLEJEYILREKNKNIDISAKVETKESTEPNEYPNGVTYPLSYND 415

```

Db      92 DGGDTFEENINVL-----K KKKKKKKNSDT-----INYNKKYFN 124

```

QY 416 IN---NALNELNSFGDLINPDYTKEPSKNIYTDNERKKFINEIKKIEKKIESDK 471

Db 125 KNKHGSSSNEHSSYD-----ENFFEAARKRKILNE-----EYKNDNEDM 167

QY 472 KSYEDRSKSLNDITKEYEKLNEIYDSKFNNIDLTFE-----KMMGKRY---SYKV 521

Db 168 MC--DMSIFND-----NNMDDSLFNKEDNNRYDEEEMKKYRKGRKRYSSDSYKD 216

QY 522 EKLTHHTFASYENSKHNLEKLTALKALYMEDYSLRNIVVEKELKYKNLISKIENEIETL 581

Db 217 DSPQMS--ERYSSSEKYSSEKYSSSN-----RQSTNLLNNIKFCNTY 268

QY 582 VENIKD-----EEQJFEKKITKD-----ENKPEKILEV--SDI-VKVQ 618

DB 269 IINKKDRSRDYEDEESREGAYGENTTEDLNEDTQEGHKNKKKEILMNLINDINIKKN 320

QY 619 VQRVLELMN-----KIDELAKTQLILKN-VELKHNTHV-----PNSYKQENKQEPYYL 664

DD 329 DERDFFMDRNFAGAKKEIDIANNNQQVMKNNENIDVINDKNF.INIDDKCPSGYF 389

000 146881CABAYTMYKVSJLHNEKNNT-KTEQ3DANSEF3IEGDL10VATIRFGQJAGS /Z

000
000

[illegible][illegible]

Q: 702 I N T E R V I E W W I T H S H E R M A N . P A R T I V O U Y A T M E M B E R S C O O R D I N A T O R I N T A N T A N T I N T E R N A T I O N A L

449 NKNSFYNN---IENRKIININNIYOKVKI-----TISPKSC-----403

0v 841 DSI NNSI SOL EMEIYKEMVCNI YKI KNDKTKNI IFEAKVSTSVKTI SSSSMODISI T 900

Db 483 DHLNVEKKL--NVTESLIC-IFDIODGDNNNNTNDP----- 517

QV 901 PODKPEVSANDDTSHSTNINNSLKLFENILSLGKNKNIVOELIGOKSSSENFYEKTLKDSD 960

```

Db      518  --DDNDDGGHHDDCGHHDDHVGNNMK-----GDK-----LDIK 546

```

Qy 961 TFYNESFTNFVSKADDINSL---NDESKRKKLEEDINKLKTQLQSFDFLYNKYKLLER 1017

PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AA157837.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PT
XX
XX Example 3; SEQ ID NO 1826; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 1780 AA;
XX
XX Query Match 4.5%; Score 375; DB 22; Length 1780;
XX Best Local Similarity 18.9%; Pred. NO. 3.le-08;
XX Matches 291; Conservative 267; Mismatches 563; Indels 422; Gaps 60;
XX
QY 130 HVRNYL-----LTIKELKYPOLFDTNH-----MLTLCNDIHFKYLDIG 170
DB 196 HRSREYLRSSQEKETASKSALLRQIKVETVHNDSDDTLYGSUTNSLITSEFEESIKD 255
QY 171 YEEIN-ELLYKLN-----YFOLLRAKLDVANDYQIIPFNKIRANELDVLKLVFGY 224
DB 256 YEQANLNMANSTKFSVWVSFFIYIYDLF-----VPVSSKQKRMKRLSODVKGY 309
QY 225 R--KPLDNIDKONVGRMEDY-----IKKNKKTNIENELIEESKRTIDKNKNATKEEEK 276
DB 310 SFIKDLQWIOVSDSK-EAYRLKLGKIQHVSVAFTKLNNASSRSHSFTFTIILQIEDSEMS 368
QY 277 KLYQAOYDLSTYN-----KOLEAHNL-----ISVLEKRIDLTKNE----- 313
DB 369 RVIRVS-ELSLCDLAGSERTMTKQNEGERLRETNGTINTSLTLGRCINVLKNSKSKFQQ 427
QY 314 -----NIKELL-----DKINEIKNPPPPANSNGTNTLLD 342
DB 428 HVPFRESKLTHTYFQSFNGCKGICMIVNISQCYLAYDETFLNVLKFSATAQKVCVPDTLNS 487
QY 343 KKKKEEKEKEIKETIKFN--IDSFTPTPLEYLYLRKNKNIDISAKVETRESTEP 400
DB 488 SOEKLFGPVKSSQDVSLSNSKILNVKTRATISWNSLEDLMEDEDLVEELENAEETQ- 546
QY 401 NEYPNGVTPPLSYNDINNALNELNSFGDLNPFDTKPSKNI--YTDNERKKFTNEIKE 458
DB 547 -----NVETKLLDELDLTKLEENKAF-----ISHEEKRKLLDLIEDLKKKLINEKKE 593

QY 459 KIKIEKKKIESDKSYEDRSKSLNDITREY-----EKLNEIYDSKFNNDIL 506
DB 594 KLTLEFKIRE-----EVTQFTQYWAQREADFKETLLQEREILEENAEURL 639
QY 507 TNFEKMMGKRYSKYVEKLTHTHTNTFASYSKHNHLEKALKYMEDYSLRNIVVEK-ELK 565
DB 640 AIFKDLVGKCD-----TREEAAKDICATKVETBEATACTACLELKFQKAEIAKTGELI 692
QY 566 YKKNLISKIENIEITLVENIKKDEQLEFKITTKDENKPDKEILEVSDIVKVQVKVLLM 625
DB 693 KTKELKKRENESSLIQLETSNKKI---ITQNG-----RIKELINIIDQREDTI 740
QY 626 NKIDELKKTQTLILKNVELKHNHVPNSYKQENKQBPYLIIVLKKKEIDKLKVPMPK-VESL 684
DB 741 NEFQNLKS-----HMENTFKCNDKADTSSLIINNNKLCINETVEVPKDSKSK 786
QY 685 INEEKKNTKTEGQSDNSEPSTEGETTGATTKPGQOAGSALBGSDVQAQKQAOQAPPV 744
DB 787 ICSEKRVNENELQODEPPAKKGI-----HVSASAI-----TEDOKKSEYR 828
QY 745 PVPPEAKAQVPTPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHYILVSHSTWNEKITL 804
DB 829 P-----NIAEIEDIRVLQE-----NNEGL 847
QY 805 KOYKITKEEESKSSCDPLD---LFNIQNNIPVMYSMFDLSLNNLSQLFMIEYKEMVC 861
DB 848 RAFLTTIENELKNEKEKAELNKQIVHFQOEL---SLSEKKNLTLSKEVQOIQSNYDIA 903
QY 862 ---NLYKLKONDKTKNLEEAKKYSTSVKTLSSSMQPLSLTPODKPEVSANDDTSHST 917
DB 904 TAEHVQKSKNOEQEIKMKLSNEIETATRSITNNVSO-IKLMHTYKIDELRTLDSVSQIS 962
QY 918 NL-----NNSLKLFEINLSLGNKNYIQLIGKQKSENFY----- 952
DB 963 NIDLLNRLDLSGSEBDNLPNTQLDGLNDYLVSKQVREY--IQPNRENSFHSSIEAI 1020
QY 953 -----EKILKSDSTFYN-----ESFTNFVKRADDINSLSNDESKRKKLEEDINKLKK 999
DB 1021 WEECKEIVKASSKSHQIELEEQIEKLAQAEVKYKIDENNRLE--KEHKNODDLKKE 1078
QY 1000 TL--QLSFDLYNK---YKLELRFDKKKTVCK--YKMOIKKL-TLLKEOLESKL 1046
DB 1079 TLIIQKLEELOEKQKNTLDVQIOHVVEGRALSELTOGVTYKAKIKETILTIETOKVERS 1138
QY 1047 NSLNNPKHVLFQNSVFFN-----KKKEAEIAETENTLE 1079
DB 1139 HSAKLEQDILEKESILKLERNLKEFBQHLQDSVKNTRKDLNVKELKLE-EITQLTNLQ 1197
QY 1080 NTKILLKHYKGLVKYNGESSPLK-TLSEESIQT-----BDNYASL-----E 1120
DB 1198 DMKHLQLKEEBEETNRQETEKLEELSASSARTQNLKADLQKBEVDYADLKEKLTDAKK 1257
QY 1121 NPKVLSKLEGLKKNLNLKLEKKLSYLSGLHLHIAELKEVIKNKNYTGNSPENNTDNN 1180
DB 1258 QIKQVQKEYSVMRDEBKLLRIKINELEKKNQCSOELD--MKORTIQQLKEQLNNQKVEE 1315
QY 1181 ALESYKKFLPEGT-----DVATVVSSESGDTLQSQPKKPASTHV-----GAE 1223
DB 1316 ALQOERACKDLNVKEIITEDMRMTLEEQEQTVQEDQVLEAKLEEVERLATELEKWKKE 1375
QY 1224 SNTITTSQNVQDVEDDVIIVPIFGESEEDYDDLQGVVGTGEAVTPPSVIDNLSKIENEYEV 1283
DB 1376 CNDLETKNQRSN-----KEHENNTDVLGKLTN-----LQDELOESEQKYNA 1417
QY 1284 LYLKPLAGVYRSLKQLENNVMTF-----NVNKKDIL-----NSRPNKRNFKNVLES 1331
DB 1418 -----DRKKLEEKMLMTQAKEAENIRNKKMKYADREFFKQONEMELITLA 1466
QY 1332 DLIPYKDLTSSNYVVKDPYKFLNKEKRFSLSYN-YIKDSITDITINFANDVLGYKILS 1390
DB 1467 QLTE-KD-----SDLOKW--REERDOLVAALQLKALISSNVQKDNIEOLKRIIS 1515


```
Qy* 433 #DYTKPSKNTYTDERKKFINEKIKIEKKIESDKSYEDRSKSLNDITKEYEKL 492
Db 163 FNLEK-LKKNLGKSGNSNENILN-----DSQKIENDKQN-TNLSKEKNS-----ENIL 207
Qy 493 NEIYDSKFNNDIDLTNEK-----MMGK---RYSYKVEK-----LTHH 527
Db 208 KTPDNSKYSNHNNTTSKKISSNSOKSELSPPSQTIGKIYRPSYLIKKELYEILDDI 267
Qy 528 NTFASYENSKHNLEKTKALKYMEDYSLRNIVWEKELKYKN-LISKIENEIETLVENIK 586
Db 268 NT-GRVTLGKNRLKELIKKGLSNFKQVKNELIENSKKEASNLLTLTKKDIEPNLINIP 326
Qy 587 KD--EEOLEFKKITYDENKPKILEVSDIVKOVQVLLMNKIDELKKTOLILKNVE-- 642
Db 327 KDPYKKEIFQ--LKDQKPK-----QVLEDLKSQVHSIKPID-LENTKSROQAIKDLNEF 378
Qy 643 LKHNIHVPNSYKQENKQOPPYLIVLKKEIDKLKVPMPKVESLINEEKKNIKTEGOS--DN 700
Db 379 LKNN---PND-AQASK-----TLAQANKIOHLEDLKSQVHSIKPIDLENTKSROQAIKDL 429
Qy 701 SEPTGEITQOATTKPCQQAQSALEGSQVQAQEQQAQPPVPVPEAKAQVPTPPA 760
Db 430 NEFLKNPNDAQA-SKTLAQANKIOHLEDLKSQVHSIK-----PIDLENTK----- 474
Qy 761 PVNKNTEVSKLDYLEKLEPLNTSYCHKYILVSHSTMNEKILKQYKITKEE--SKLS 818
Db 475 -----SRQAIDKIDNEFLKNN-----PNDQAASKTLAQANKIOHLEDLKSQVH 517
Qy 819 SCDPLDLLFNQNNIPVMSYMFDSLNLSOLFMEIYEKEMVCNLYKLKND-KIKNLE 877
Db 518 SIKPIDL--EN-----TKSRQQAIDKIDNE-----FXKNPNDAQASKTLA 555
Qy 878 EAKVSTSVKTLSS--SMOPLSLTPQDKPEVSANDDTSHSTNLNLSLKFENILSLGKN 935
Db 556 QANKIQ-HLEDLKSQVHSIKPIDLENTKSROQAIKDLNEFXKNPNDAQASKTLAQANKI 614
Qy 936 KNIIQELIGQKSSNFYEKILKSDSTFYNESFTNFVSKKADDIN--SLNDESKKKLEED 993
Db 615 QH-----LED-----LKSQVHSIKPIDLENTKSROQAIK 644
Qy 994 INKLKKT----LOLSFDLYNKYKLEKRLFDKKTQVKKYKMQIKKLTLLKEOLESKNSL 1049
Db 645 LNEFLKNPNDAQASKTLAQAN--KIOHLEDLKSQVHSIKPIDLENTKSROQAIKDLNEF 702
Qy 1050 --NNPKHVQLNFSVFNNKKAETAEENTLEN---TKILLKHYK-GLVKYNGESSPL 1102
Db 703 XKNPNDAQASKTLAQAYENNGDLLKAENAYEKIUKLTQEDHYKGLIRF-----KL 756
Qy 1103 KTLSESIQTDNVTASLENFKVLSKLEKGLKDNLNLEKKLSYLSGSLHHLIAELKEVIK 1162
Db 757 KKYEH-----SIESFDQTIKLDPKHK-----KALHNKGIALMMLNK 792
Qy 1163 NKNYTGNPSNNTDVNNALESYKFFLPEGTDVATVVSSESDTLEOSOPKPPASTHVA 1222
Db 793 NK-----KAIESFEKAIOIDKNYGTAYQKG-----IAE 821
Qy 1223 ESN-----TITTSQNVDDVDVITVPIFGESEEDYDLGQWVTGEAVTPSVIDNIL 1274
Db 822 EKNGDMQQAFAFNKAYNLD-----KNPNY-----ALKAGIVSNL 857
Qy 1275 SKIENEYEVLYKPLAGYRSLKQLENVNMTFNNVKD-----ILNSRFNKRNFKNVL 1329
Db 858 GNF-----KOSEYLNFFNANAKKPEIAIYNLSIAKFEN--NKL 895
Qy 1330 ESDLIPYKDLTSSNYYVVKDPKFLNKEKRDKFLSYNYIKDSIDTDINFANDVLGYKIL 1389
Db 896 E-----ESLETINKAID-----LNPEK-----SEYLYLKASINLKNYQNAISLSLV 939
Qy 1390 SEKYSKDLSDIKKYN-----DKQGE-----NEKYLPLANNIETLY---KTVN 1429
Db 940 IEK---NPENTSAVINLAKAYEKGSKNSQAISTLEKILNKNNKLANLGLIYKKEKNYQ 996
Qy 1430 DKIDLFIHLEAKVLYNTYE--KSNVEVKIKELNLYKTQDKLADF---KKNK-----NFVG 1481
```

```
Db 997 KAIEIF---EKALINSIEAKYNLATTLIEINDNTRAKDLLREYTKLKPNPEALHALG 1052
Qy 1482 IADLSTDYNNHN-----LLTKF 1498
Db 1053 I-----IETNENNNDOTLRELKIF 1072

RESULT 54
AAY19934
ID AAY19934 standard; Protein; 1119 AA.
XX AAY19934;
DT 19-JUL-1999 (first entry)
DE B. burgdorferi antigenic protein, f742.aa.
XX Antigenic protein; vaccine; Lyme disease; infection; detection.
OS Borrelia burgdorferi.
PN WO9859071-A1.
XX 30-DEC-1998.
XX 18-JUN-1998; 98WO-US127118.
XX 03-SEP-1997; 97US-0057483.
XX 20-JUN-1997; 97US-0050339.
XX 22-JUL-1997; 97US-0053344.
XX 22-JUL-1997; 97US-0053377.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (MEDI-) MEDIMMUNE INC.
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
WPI; 1999-189980/16.
DR N-PSDB; AAX61631.
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX Claim 12; Page 124-125; 275pp; English.
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX SQ Sequence 1119 AA;
```

Query Match 4.48; Score 369; DB 20; Length 1119;
Best Local Similarity 21.38; Pred. No. 3.3e-08;
Matches 286; Conservative 208; Mismatches 472; Indels 378; Gaps 68;

```
Qy 261 KTDKNKNATKEEKKLYQAQYDLSIYNKOLEEAHNLSVLEKRIDTFLKKNENIKELLD 320
Db 33 KLNDKNEIMLNEVKN-----SVIDRNYKKAYSVAKLLQ---DKYPONEDIAMLTN 80
Qy 321 KINIKNPPPPANS-----GNTPTNTLLDNKKTEEHEKEIKE---TAKTIKFNIDSLFTDPL 373
Db 81 TLAIANSPPESKDLQDSANQILDKIKGODNTKTNVNFDAFNRYIKDSTITE-- 138
Qy 374 ELEYLREKKNKIDISAKVETKESTPENEY-PNGVTYPLSYNDINNALNELNSFGDLNLP 432
Db 139 --NYSRDNDDVGIEDIESEFKKIKPEIKPN--TNPKEEDQIIQSPNPKSLVNDQKNL 194
```


Qy 1484 DLSTDYHNHLLTKFLSTGVMFENLAKTVLSNLLDGNLQGLNISQ-----HOCVKKQC 1537
Db 2247 KTNKORNNKELKRSISLPL---NLKRTVVKII---NLKKNLNKNIIDAINDLKGT 2300
Qy 1538 P-----QNSGCRHLDERECKC-----LNNYKOBGDKCVEN 1569
Db 2301 PYEHYTHSNFWIYSSSDSYNCSDNYLNNVDHGSKEFDN 2343
RESULT 57
ABB63519
ID ABB63519 standard; Protein; 2346 AA.
XX AC
XX ABB63519;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 17349.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmacetical.
XX
XX Drosophila melanogaster.
XX OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX PF
XX 23-MAR-2000; 2000US-191637P.
XX PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR
XX N-PSDB; ABL07622.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 17349; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2346 AA;
XX
XX Query Match 4.3%; Score 363; DB 22; Length 2346;
XX Best Local Similarity 20.0%; Pred. No. 1.5e-07;
XX Matches 278; Conservative 240; Mismatches 525; Indels 344; Gaps 57;
Qy 171 YEEINELLYKLNFYDILLRAKLDVNCANDYQIFPNLKIRANELDVLLKLVFGYRKPLDN 230
Db 395 YAKSESELEMRNCEIQKQLKSIIAFISAP-----ILEKQNSDYQK-MKE 442
Qy 231 IKDNWGMEDYIKKNKKTIIENINELIESKTTIDKNKNATKEEKKKLYQAQYDLSIY-- 288
Db 443 TNSELLREHDELLQNLCLC---RELERALSTLHNQ---ENKKLKQTHDLSROVC 494

Qy 289 -----NKQLEEAHNLISVLEKRIDTLTKNENIKELDKINEIKNP 328
Db 495 MLDLNCIRAGVHVRIOPTROLPTSESLIS-----DNLVTFSSIEELVD-----RNT 543
Qy 329 PPANSGNTPNTLLDKNKIEHEKEIKIAKTIFKNIDSLFTDPLLEYYEYREKN----- 383
Db 544 YLLNMSRELTELLEASEK--NODKMLLEQSKNHIRKLDAREFA--ELEDLLTLQKNNTVT 598
Qy 384 -----KNIDISAKVETKEST---EPNEYVNGVTYPLSYNDINN----- 418
Db 599 LLSKCDRYKKLYFAAQKKGQNTVDLDDSNLEPNDSALDSEOPANFEESRKKLEKRVQ 658
Qy 419 ----ALNELNSFGDLINPFD-YTKEPSKNIYTDNER----KKFINEI----- 456
Db 659 LEQOLEGEVKKYASLKENYDIYTSKRNKDALAOEQFDSMRKEVRELTSNCKLMTTTF 718
Qy 457 -KEKIKIEKKKIESDKSYEDRSKSLNDITKEYEKLLEIYDSRFNNNIDITLNFPEKMGK 515
Db 719 QKEQIELLHKNI----GTYKQQTTLBERTKNYEKTI-----IKHEQTVHLIKDEMMAH 769
Qy 516 R-----YSYKVEKLTTHN-----TFASYENSKHNLEKL-- 543
Db 770 RKHAAADAEAOQLRQENRILDRITSSRLQIERETHREOQSOSLLNSLEFIKTNLERSEM 829
Qy 544 --TKALKYMEDYSLRNIVVEK-----ELKYKNLISKIENIETLVENIKKDEEQLFEK- 595
Db 830 EGRQRLQRLDDTVRELAARHFQEEBEKFRSINEFKROAETAIK-LKDEEKQLADKW 888
Qy 596 --KITKDENKPDKEILEVSDIVKVOVKVLLMNKIDEL-KKTQLILKKNVLEKHNTHVPS 652
Db 889 QAEELTSVREELAERK-----NKVNELSKLQEVLTPTLNDNPITAANK 931
Qy 653 YKOENKQEPYLIIVLKEIDKLVFMPKVESLINEEKNIKTEG-----QSDNSEPT 705
Db 932 RARE-----FELKLDQATV---BIESLTRELAKT-REHGEQFYKMSQSAESEIKR 977
Qy 706 EGEITQATTKPGQOAG-----SALEGDSYQAQAEQ-----KQAPQ 742
Db 978 LHELHGLVAKQEEIEKKLRSEAEKTRISDLAEAEAMLSNVTEQSKTVNQSGQLKSAQD 1037
Qy 743 PVP---VPVPEAKAQPPTPPAPVNVKNVTENVSKLDYLEKLYEFLNTSYCHYIYLVSHSTM 799
Db 1038 DLKSLLEKLTAEANTIRT-----LRSENTS-----LVESLNAAEVYKANGMTQHSAD 1084
Qy 800 NEKILKOVKITKEESKLSLSCDPLDLFNIGNIIPVMTSMFDSL--NNSLSQSLFPEIYEK 857
Db 1085 IQELTRY----KAEFFKAN-----DELNLQASGRESLQAAAYDELLRSNAEQKLLDKERE 1135
Qy 858 E---MVCNLYKLKDNDKIKNLEBAKVKVTSVKTLSSSMQP---LSLTPQDKPE-VSAN 910
Db 1136 ESEKRVADLHALN-----SNLHDQTEALASKLAVLASQSNPNSSLNESAMDGQSLNAS 1190
Qy 911 DDTSHSTNLNNSLKLFIENILSLGKNKNIYQELIGOKSSEN-----FYEKILKDSOTF 962
Db 1191 GLTAAEGERNNE-QLLKTIKFLRKEKDLFAAKLDILKAENARLISEHAIOQKKVDELNGY 1249
Qy 963 YNESFTNFVKSKADDIINSLNDESKRKKLEEDINKKTKQLSFDLYKNYKIKLERLFPDK 1022
Db 1250 LNQ---ERAKSOTDVVSANKHEEVLRKI-ETLNAITDSNRILREARNALTLRVAELTDRI 1305
Qy 1023 KTVGYKMQIKKLLTLKLEQESKLSLNPNKHVIONFVSFFNKKKEAEIATENTLENTK 1082
Db 1306 SSVEK---ELPPLQCSNKELTSKTEEINVENTSLRTEAIKWRQRANALVEKSNRNPFEK 1362
Qy 1083 ILLKHVKLVKYYNGESSPLKTLSEE----- 1108
Db 1363 RLQAREHLAKLLTAEKELNKKQSDVELTVLKORMNTEIPLMNKOMQILDEARKKQVDFT 1422
Qy 1109 -----STQTDNVDYASLENFKVLSKLECKLDNLEKLLKLSYLSGLHLHIAELKEVIK 1162
Db 1423 NLKQNTFQOTOD-TWELKN-RLLOKEEBELLKANELETKDTKIADKETKEL--QURKLA 1478
Qy 1163 NKN--YTGNSPSENNTDVNNALESYKKFLPE-GTDVATVVSSESGSDTLEQSQPKKPASTH 1219

Db 1479 RYKDFVIGLSQGGGTESAAELEKRVSELEEVNQLRALDKDHEKITKCEDEVKK---R 1534
QY 1220 VGEASTITTSQNVDDVIVPIFGSEEDYDGLQGVTFGEAVTPSVIDNILSKIN 1279
Db 1535 TPEPDTSAIROEYKAKDLVLVDLTVAFT---DLVNOETTFACKSSYDETARLEK 1589
QY 1280 EYEVLYKPLAGYRSLKKQL--ENNVMPTFNVN-VKDIILNSRFNKNFNKVNLESILPI 1336
Db 1590 E-----LQENIAANKDINQRLTRENESLHMRINQLTRQLGSOOSTKPSSTSSVAEKGNI 1644
QY 1337 KDLTSSN 1343
Db 1645 SSPRTAN 1651
RESULT 58
AAB18205
ID AAB18205 standard; Protein; 1712 AA.
AC AAB18205;
XX 07-NOV-2000 (first entry)
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:62.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX Plasmodium falciparum.
OS WO200025728-A2.
XX 11-MAY-2000.
XX 05-NOV-1999; 99WO-US26796.
XX 05-NOV-1998; 98US-0107131.
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 141-146; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA/0078 to AAA/0287 and AAB18144 to AAB18352 represent nucleotide

CC and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
SQ Sequence 1712 AA;
Query Match 4.3%; Score 362.5; DB 21; Length 1712;
Best Local Similarity 20.0%; Pred. No. 1.le-07;
Matches 317; Conservative 262; Mismatches 574; Indels 429; Gaps .81;
QY 204 PF---NLKTRANELDVLKLVFGYRKPLDNIKDNYGKMEDYIKKNKKTINELIEESK 260
Db 20 PFERNLNDKGHLKELKRL---ERVDEPLYNNDVKI-----PNKKEIYNNIKSNIIQ 71
QY 261 KTDKNKNAKTEE-----KKLYQAOYDLSIYNKOLEEAHNLISVLE--- 303
Db 72 VRVQNNEEKKKEBANYTCVNNKYVTLKNKVHVNYK---VNSNINKI-KIVPIIKCSN 127
QY 304 -----KRDTLAKN-ENIKELLDKINEIKNPPANGNTPNTLLDKNKKIEBEKEIKEI 357
Db 128 YKIKNNPISHLKSNEYENFKLSNFSNIKN-GCSHKDNVINETMDQHKSEQLNNDNIKKL 186
QY 358 -----AKT-IFKNIDSLFDPLELYLREKNKNIDISAKVETKESTEPNEY 403
Db 187 LDYDIFREDTITKTNTNISYNNKMSFKDNEENINYM--DNNNIKSSSSYCSYKNIKO- 243
QY 404 PNGVTYPLSYNDINNA-----LNELNS-----FGDLINPFDTYKPSKNIYTDN 447
Db 244 -----NNVHTLKTFELEKNSHTQNEQSTPLDGLQNNHNSATKPHNNIYDNN 293
QY 448 ERKKFINEIKEK-IKIEKK--KIESDKSYEDRSLSNDITKEYEKLNLNEIYDSKFN--- 501
Db 294 --NSLVNYKSGDIDLHNKMMKIETDKNGIITLEKKKHD-----EKYYNINFLNLDNS 346
QY 502 NNIDLTNFEKMMGKRYSYK--VEKLTHHTTFASYENSKHNLEKLTAKALYMEDYSLRNIV 559
Db 347 NNVVITTCNKESYRNSTSDMINKI-----PEKMMNEKKNILK----MKNFNDVTKKKT 397
QY 560 VEKELKYKNLISKIENEIETLVENIKK-----DEEQLEFKKI-----TKDENKPDCK 607
Db 398 MAKE-----KILNSNSTI--NMKKVSYFNSKDEDLFNEKENSYKYGVRKNEQEDIN 446
QY 608 ILEVSDIVKVQVQVLLMKNKIDELAKTQL-----ILKNVELKHNIHVNSYKQENK 658
Db 447 VIK-----NNMKRNINIDNNDNINIIKNDYSKNIHI--NNKKKRD 487
QY 659 QBPYVL---IVLKKELDKLV--FMPKYVESLINEEKNKIKTEGQSDNSPSETEIGQA 713
Db 488 DPPFNNSAGLLDFDLCKRKKVLEIKNVQS--SKKKKILTNHNSNQNCHSSDNQNC 545
QY 714 TTKPQQQAGSA-----LEGDSYQAOAQEQKQAQPPVPVPVPEAKAQVPTPPAP 761
Db 546 HSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDN 593
QY 762 VNNKNTYNSKLDYLEKLYEFLNTSYICHKY ILVSHSTWNEKTLKQYKIT--KEESKSLSS 819
Db 594 MKNKSNNSK-----NKRETKSKKIKNNNNNNNNNNNN 625
QY 820 CDPL-DLLFNION---NIPVMYSMFDLSLNSLSOLFMEIYEKEMVCNLYKLKDNDIKN 874
Db 626 CDMGDQRINNNDKQNVNIONEGNFNNNKNNNDLLNVIISPNNIN-HSLSSCEKKN 684
QY 875 LLEAAKYSTVSKTSSSSMQPLSLTPQDQPEVSANDDTSHSTNL--NNSKLFLNLSIG 933
Db 685 --KEDKNNDN-KFLNSSSKMKI-----PEISTNSNEKIVNYSNDEMVLVHNLTVLN 734
QY 934 KKNKIYQELIGOKSS-----ENFYKILKDSOTFFNESTNFVKSKADDINLSDS 985
Db 735 VKE--QGVTEBESSICKRTVEVDQFY-----DSYNMRNEKITD-DNMQVEDIYNVKENI 785
QY 986 KRKKLEEDINKLKTQLQSF-----LYNKYKLEKLERLF----- 1019
Db 786 KR-TLKGOGHDDVKTNMLSEDSNYSAGLWNGNFINISNNENCLNSYDISCDEKYPNEEE 844


```
Qy 632 KKTQILKNVLEKHNIHVPNSYKQENKQEPYLIIVLKKKEIDKLVFMPKVESLINEEKN 691
Db 544 RRGQKI--GOEKH-----KKDENK-----KNKN 568
Qy 692 IKTEGOSNSEPSTGEITGOATTPGQOAGSALBGDSVQAQOQKQAQPPVPVPEA 751
Db 569 KNKNNSNNINNKHGRVITQYTDK-----IQDYCKNKESKRGNHKKMRK 615
Qy 752 KAQVTPPAPVNNKTENVSKLDYLEKYLEFN-----TSYI-CHKYILV-----S 795
Db 616 EKNLNSLLSINGKCYNKWKKNY-NKTRKPKNEGKGEKIYCYENIKILEDIKDRFFND 674
Qy 796 HSTMN-----EKIKQYKITKEESKLSCDPLDLFNI-----ONNIP 834
Db 675 HKNRNLNEENFIKEHQINGRKNKEHVNEKNKEEDFNISKENTKEGSYIIHKNRNDN 734
Qy 835 VMYSMFDLSNLSOLFMEIYEKEMVCN-LYK-LKNDKIK-----NLLEE-----878
Db 735 IKIGRYDIND-----KKEPSSNLYKCVKKNKINKSQTSLFFEFMKGKDQKH 784
Qy 879 --AKKVSTSVKTLSSSSMQPLSLTPQDKPEVSANDDTSHSTNLNSLKLFFENILSLGKNK 936
Db 785 NVIKKEDVFIKTFRTNK-SPELTAK-----ISDYKCNL-----LYTSLDRIHKNV 829
Qy 937 NIYQELI-----GOKSSENFYEKILKSDTFPYNESFTNFVKSADINDINLDESK-----986
Db 830 SIYNERIERTKHVPQKNDNI-----DIRGIYS--YNFKS-----NMNMSLSKCYHT 877
Qy 987 ----RKLEEDINKUKTLQ--LSFDLYNKYKLERLFDKKTGVKYQMOKKUTLLKE 1040
Db 878 KTCDSYNDFMKNKSKKAQNKLVSKCISKYKKAKKERKETTTRKKKIYRKEISI 937
Qy 1041 QLESKLNSLNNPKHVLQNFVSFFNKKKAETAEENTLENTKILLKHYKGLVKYVYNGESS 1100
Db 938 SFDGNVFGHENKRTKEN-----NKSKE--AYTSKRNNKIKGEEKTKRSLCSYKLR 990
Qy 1101 PLKTLSEE-----SIQTEDNYASL--ENFKVLSKLEGLKLDNLN 1137
Db 991 KMKHLCVENKMHIKKNVRQI IKKKKKNIYKTIKCLNSYKTLIDQVNVKGDDEHKLSNHVN 1050
Qy 1138 LEKKLSYLSLGLHLIAELAEKVIKNKNTGNSPENNTDVN-NALESYKFLPEGTDVA 1196
Db 1051 NKKKKKK-----NCINENDDNNDNYNDNNDNNDNNDNNDNY-----1092
Qy 1197 TVVSESGDTRLRQSPKPPASTHVGAESENTITTSQNDVDEVDVILVIPFGESEEDYDDL 1256
Db 1093 ----NDNNND-----HNDHNNNDNNDLNDHNDN-----NQREHSCEBI 1129
Qy 1257 GOVVTGEAVTPSVIDNILSKIENEYEVLYLPLAGVYRSLKQLENNVMTFNVVYK----1312
Db 1130 N-----IQNVEQKCEGEYEGKRKNKYTYNNYKINGKNEIHDYDIKSHGS 1177
Qy 1313 ----DILASRFNKRNFKN-----VLESDLIPY-----KDLTSSNYVV-----KOPY 1350
Db 1178 RYNYIFNIFIKDNKHNNDKGEKSELKCSIDPYVEKYENLENNTYEIIGLIYYGDKSQY 1237
Qy 1351 KFLNKEKR-----DKFLSSYNIKDSITDINFANDVL---GYYYKILS 1390
Db 1238 KCINNNKRVYAMKVVLKCEKNEIFVDNFIKYLFLKNNPHKNIIISYDFCNNNYICIIM 1297
Qy 1391 EYKSDLSIKKYINDKOGENEKYLPLFNLNITLYKTVDNKIDLF-----VHLEAKVLNY 1446
Db 1298 D-YCEGSTLLDYFMSLVPGSLDVY-----EIKKIMKNPIALDFPHSNNIIHRDIKLENI 1351
Qy 1447 TYEKSNEVEKIKELAYLKTIODLADFKKNNNFVGIADLSTDYNNHNLTLTFLS-TGMVF 1505
Db 1352 MFK--NKKRKKRFRNY-----EKYSGLFNH-----EISFSTSCSNLHKDLQLRGM--1398
Qy 1506 ENLAKTVLS-----NLLDGNLQGMNLSQHOC-----VKQCPQNSGCFRHLDERECKC 1555
Db 1399 DTIGKKGKGGKFIIRNLNYNEKHKN-LNIFQKNCSHLLKKNTKKN--ILSNDIQLKSPKC 1455
```

Qy 1556 LLNYKQEGD 1564

Db 1456 YIKYNNND 1464

Search completed: March 31, 2003, 07:20:42
Job time : 161 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 07:18:27 ; Search time 67 Seconds
(without alignments)
5040.466 Million cell updates/sec

Title: US-09-269-874A-3

Perfect score: 8424

Sequence: 1 MKIIFFLCFLFFIINTQCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database :

SPTREMBL_21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mmc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5283.5	62.7	1704	5	Q9T2T4
2	5162.5	61.3	1694	5	Q9NHX1
3	5162.5	61.3	1694	5	Q9T2T5
4	5107.5	60.6	1720	5	Q25922
5	3361	39.9	656	5	Q8T6A9
6	3093.5	36.7	1751	5	Q26194
7	3003.5	35.7	1087	5	Q25961
8	2951	35.0	1726	5	Q02569
9	2801	33.3	539	5	Q9TYG1
10	2798	33.2	539	5	Q25972
11	2796	33.2	539	5	Q25966
12	2795	33.2	539	5	Q25973
13	2794	33.2	539	5	Q25976
14	2791	33.1	539	5	Q25981
15	2789	33.1	539	5	Q25984
16	2788	33.1	539	5	Q25971

Q25645 plasmodium
Q9U6D4 plasmodium
Q25685 plasmodium
Q25668 plasmodium
Q25924 plasmodium
Q25923 plasmodium
Q03999 plasmodium
Q9BNG8 plasmodium
Q9NAT3 plasmodium
Q25975 plasmodium
Q25969 plasmodium
Q25977 plasmodium
Q25974 plasmodium
Q25979 plasmodium
Q25983 plasmodium
Q25967 plasmodium
Q25968 plasmodium
Q25982 plasmodium
Q25970 plasmodium
Q9TYG2 plasmodium
Q25978 plasmodium
Q25978 plasmodium
Q25865 plasmodium
Q9T2U5 plasmodium
Q9T2U2 plasmodium
Q9T2U7 plasmodium
Q9T2U8 plasmodium
Q9T2V2 plasmodium
Q9T2V2 plasmodium
Q9T2U9 plasmodium
Q9T2U1 plasmodium
Q9T2V3 plasmodium
Q9T2V1 plasmodium
Q9T2U4 plasmodium
Q9T2U3 plasmodium
Q9T2V0 plasmodium
Q9T2V0 plasmodium
Q9T2V6 plasmodium
Q9T2V5 plasmodium
Q9T2U6 plasmodium
Q9T2V4 plasmodium
Q9T2V8 plasmodium
Q9T2T9 plasmodium
Q25760 plasmodium
Q9T2W5 plasmodium
Q9T2W1 plasmodium
Q9T2V9 plasmodium
Q9T2W3 plasmodium
Q9T2W7 plasmodium
Q9T2U0 plasmodium
Q9T2T7 plasmodium
Q9T2W8 plasmodium
Q9T2W4 plasmodium
Q9T2W6 plasmodium
Q9T2W2 plasmodium
Q9T2V8 plasmodium
Q9T2V7 plasmodium
Q9T2W0 plasmodium
Q25724 plasmodium
Q25725 plasmodium
Q25721 plasmodium
Q43997 plasmodium
Q25726 plasmodium
Q9NCM9 plasmodium
Q25722 plasmodium
Q25723 plasmodium
Q9NCM1 plasmodium
Q9NCM4 plasmodium
Q9NCM2 plasmodium
Q9NCN2 plasmodium
Q43995 plasmodium
Q25719 plasmodium
Q25720 plasmodium
Q25727 plasmodium

90	1173	13.9	372	5	Q25717	Q25717 plasmodium
91	1171	13.9	219	5	Q9UB87	Q9UB87 plasmodium
92	1170.5	13.9	373	5	Q43996	Q43996 plasmodium
93	1170.5	13.9	599	5	Q9NCM5	Q9NCM5 plasmodium
94	1168.5	13.9	373	5	Q25728	Q25728 plasmodium
95	1168	13.9	372	5	Q25718	Q25718 plasmodium
96	1165	13.8	239	5	Q9NGK6	Q9NGK6 plasmodium
97	1164.5	13.8	597	5	Q9NCN3	Q9NCN3 plasmodium
98	1164.5	13.8	597	5	Q9NCM3	Q9NCM3 plasmodium
99	1160.5	13.8	597	5	Q9NCM6	Q9NCM6 plasmodium
100	1159.5	13.8	595	5	Q9NCM0	Q9NCM0 plasmodium
101	1158.5	13.8	597	5	Q9N657	Q9N657 plasmodium
102	1157.5	13.7	597	5	Q9NCN1	Q9NCN1 plasmodium
103	1150.5	13.7	599	5	Q9NCM7	Q9NCM7 plasmodium
104	1147.5	13.6	613	5	Q9NCN0	Q9NCN0 plasmodium
105	1147	13.6	616	5	Q9NCM8	Q9NCM8 plasmodium
106	1105	13.1	213	5	Q9TY76	Q9TY76 plasmodium
107	1105	13.1	213	5	Q9TY75	Q9TY75 plasmodium
108	1105	13.1	213	5	Q9TY73	Q9TY73 plasmodium
109	1105	13.1	213	5	Q9TY72	Q9TY72 plasmodium
110	1105	13.1	213	5	Q9TY70	Q9TY70 plasmodium
111	1099	13.0	213	5	Q25928	Q25928 plasmodium
112	1099	13.0	213	5	Q25939	Q25939 plasmodium
113	1034	12.3	211	5	Q25707	Q25707 plasmodium
114	1021.5	12.1	636	5	Q26182	Q26182 plasmodium
115	992.5	11.8	214	5	Q25927	Q25927 plasmodium
116	992.5	11.8	214	5	Q25930	Q25930 plasmodium
117	977.5	11.6	214	5	Q25942	Q25942 plasmodium
118	973.5	11.6	214	5	Q25936	Q25936 plasmodium
119	968.5	11.5	214	5	Q25943	Q25943 plasmodium
120	954.5	11.3	214	5	Q25926	Q25926 plasmodium
121	954.5	11.3	214	5	Q25931	Q25931 plasmodium
122	954.5	11.3	214	5	Q25934	Q25934 plasmodium
123	954.5	11.3	214	5	Q25937	Q25937 plasmodium
124	954.5	11.3	214	5	Q25938	Q25938 plasmodium
125	939.5	11.2	214	5	Q25929	Q25929 plasmodium
126	939.5	11.2	214	5	Q25932	Q25932 plasmodium
127	939.5	11.2	214	5	Q25933	Q25933 plasmodium
128	909	10.8	215	5	Q25935	Q25935 plasmodium
129	909	10.8	215	5	Q25940	Q25940 plasmodium
130	827.5	9.8	218	5	Q9TVG8	Q9TVG8 plasmodium
131	689.5	8.2	379	5	Q25659	Q25659 plasmodium
132	686.5	8.1	144	5	Q25992	Q25992 plasmodium
133	659.5	7.8	138	5	Q25960	Q25960 plasmodium
134	656	7.8	121	5	Q9UAI7	Q9UAI7 plasmodium
135	656	7.8	121	5	Q9TYE4	Q9TYE4 plasmodium
136	655	7.8	121	5	Q9UAI6	Q9UAI6 plasmodium
137	649	7.7	119	5	Q9TYE7	Q9TYE7 plasmodium
138	645	7.7	121	5	Q9UAI8	Q9UAI8 plasmodium
139	642	7.6	327	5	Q26043	Q26043 plasmodium
140	637	7.6	121	5	Q9TYE6	Q9TYE6 plasmodium
141	637	7.6	312	5	Q00796	Q00796 plasmodium
142	634.5	7.5	315	5	Q00794	Q00794 plasmodium
143	634.5	7.5	315	5	Q00798	Q00798 plasmodium
144	634.5	7.5	315	5	Q00803	Q00803 plasmodium
145	634	7.5	116	5	Q9TYE3	Q9TYE3 plasmodium
146	633.5	7.5	315	5	Q00799	Q00799 plasmodium
147	631	7.5	312	5	Q00797	Q00797 plasmodium
148	631	7.5	328	5	Q9TYF2	Q9TYF2 plasmodium
149	630	7.5	328	5	Q9TYF0	Q9TYF0 plasmodium
150	628.5	7.5	315	5	Q00802	Q00802 plasmodium
151	628	7.5	338	5	Q9TYE9	Q9TYE9 plasmodium
152	628	7.5	338	5	Q9TYF1	Q9TYF1 plasmodium
153	628	7.5	338	5	Q9TYF3	Q9TYF3 plasmodium
154	626.5	7.4	339	5	Q00792	Q00792 plasmodium
155	625	7.4	304	5	Q00801	Q00801 plasmodium
156	625	7.4	338	5	Q9TYF4	Q9TYF4 plasmodium
157	624	7.4	338	5	Q00791	Q00791 plasmodium
158	624	7.4	338	5	Q00793	Q00793 plasmodium
159	623	7.4	338	5	Q9TYE8	Q9TYE8 plasmodium
160	619.5	7.4	328	5	Q9GSQ9	Q9GSQ9 plasmodium
161	617.5	7.3	339	5	Q00800	Q00800 plasmodium
162	616.5	7.3	333	5	Q00795	Q00795 plasmodium
163	610.5	7.2	2771	5	Q26216	Q26216 plasmodium
164	606.5	7.2	1939	5	Q25662	Q25662 plasmodium
165	568.5	6.7	2269	5	Q26223	Q26223 plasmodium
166	559.5	6.6	1979	5	Q96133	Q96133 plasmodium
167	558.5	6.6	2747	5	Q9BJX9	Q9BJX9 plasmodium
168	558	6.6	2867	5	Q9N2M3	Q9N2M3 plasmodium
169	556	6.6	3130	5	Q9BK46	Q9BK46 plasmodium
170	554	6.6	3254	5	Q9BK45	Q9BK45 plasmodium
171	553	6.6	96	5	Q9TYE5	Q9TYE5 plasmodium
172	553	6.6	2752	5	Q9BJY0	Q9BJY0 plasmodium
173	539	6.4	105	5	Q9TY71	Q9TY71 plasmodium
174	528	6.3	95	5	Q964N2	Q964N2 plasmodium
175	525	6.2	95	5	Q964N3	Q964N3 plasmodium
176	524	6.2	105	5	Q25941	Q25941 plasmodium
177	521.5	6.2	3394	5	Q77384	Q77384 plasmodium
178	518.5	6.2	308	5	Q02560	Q02560 plasmodium
179	517.5	6.1	308	5	Q02619	Q02619 plasmodium
180	516.5	6.1	308	5	Q02620	Q02620 plasmodium

ALIGNMENTS

RESULT 1

Q9TZT4 PRELIMINARY; PRT: 1704 AA.

AC Q9TZT4; TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Merozoite surface protein 1.

GN MSPI.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

NCBI_TaxID=5833;

[1] SEQUENCE FROM N.A.

RC STRAIN=HN2;

RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;

RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from Hainan, China.";

RL	Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR	EMBL; AF062349; AAC72885.1; -.
DR	InterPro; IPR000561; EGF-like.
DR	Pfam; PF00008; EGF; 1.
SQ	SEQUENCE 1704 AA; 193763 MW; 385526D0DA56FDID CRC64;
Query Match	62.7%; Score 5283.5; DB 5; Length 1704;
Best Local Similarity	62.1%; Pred. No. 2.4e-157;
Matches 1075; Conservative 219; Mismatches 317; Indels 121; Gaps 26;	
QY 1	MKTIFFLCFLFFIINTQCVTHESYQELVKKLEDAVLITGYSLFQKEMVINEGTS 60
Db 1	MKTIFFLCFLFFIINTQCVTHESYQELVKKLEDAVLITGYSLFQKEMVINEGTS 60
QY 61	AVTSTPGSK---GSVASGGSGSVASGGSGSVASGGSGSVASGGSGNRRTNFDNS 117
Db 61	AVTSTPGSKGGSGSVASGGSGSVASGGSGVA---SVASGGSGNRRTNFDNS 117
QY 118	SDSDAKSYADLKHVRVNYLLTKELKYPOLFDLTNHLTLCDNIHGFYKYLIDGYEINEL 177
Db 118	SDSDAKSYADLKHVRVNYLLTKELKYPOLFDLTNHLTLCDNIHGFYKYLIDGYEINEL 177
QY 178	LYKLNFFDILLRAKLDVNCANDYCOIPFNLKIRANELDVLKLVFGYRKPDLNKNVKG 237
Db 178	LYKLNFFDILLRAKLDVNCANDYCOIPFNLKIRANELDVLKLVFGYRKPDLNKNVKG 237
QY 238	MEDYIKNNKTTENINELIEESKKTIDKNKNATKEEKKLYOAOYDLSTYNNKQLEAHN 297
Db 238	MEDYIKNNKTTENINELIEESKKTIDKNKNATKEEKKLYOAOYDLSTYNNKQLEAHN 297
QY 298	LTSVLEKRDITLKKNNIKELDKINEKNPPANSNGTNTLLDNKNKTEEHEKEIKEI 357

Db 298 LISVLEKRIDTLKKNENIKLLDKINEIKNPPANGSNTPTNLLDKNNKIBEHKEIKEI 357
Qy 358 AKTKFNIDSFTDPLEYLYLREKNKNDISAKVE---TKESTEPN-EYPNGVYPLSYN 414
Db 358 AKTKFNIDSFTDPLEYLYLREKNKNDISAKVE---TKESTEPN-EYPNGVYPLSYN 414
Qy 415 DINNAL---NELNGFGLINFDYTKPSKNIYTDN-ERKKFPIBIEKIKIEKKI--- 467
Db 418 DIHNSLAANDKNSYGLMNP-DTKEKINEKIITDNKERKIFINNIKKOIDLEKNIHT 476
Qy 468 -ESDKSYEDRSKSLNDITKEYEKLLNIEYDSKNNNIDLTFEKMKGRSYKVEKLT 526
Db 477 KEONKKLLDEYKS---KKDYEELEKFEYEMKNNNDKVDQVIFISARYTYNVEK 532
Qy 527 HNTFASYSKHNLEKLTALKYKMEYSLRNIVVEKEKLYKLNLSKISKEIETLVENIK 586
Db 533 NKKFSSSNYSYVQKAKALSYLEDSLKIGSEKDNHYTYLTKGLEADIKKLTIEIK 592
Qy 587 KDEOLFEEK---ITKDNKDPDEKILEVSDIVKQVQVQLVLMNKIDELKTKTOLILKNVEL 643
Db 593 SSENKILEKFKGLTHSANAS---LEVSIVKLVQVQVLLIKTKIEDLTKLTKLNAQL 648
Qy 644 KHNTHVPSYKOEKQEPYIIVLVKKEIDKLVKMPKVESLINEEKNKIKTEGOSDNSEP 703
Db 649 KDSHVPNIYKPNKPEPYIIVLVKKEIDKLVKMPKVESLINEEKNKIKTEGOSDNSEP 703
Qy 704 STEGEITQATTKPGQQAAGSALGDSVQAQAEQKQAO-----PPVPVPVPEAKAQVPT 757
Db 700 ITQPLVAASETDEGGHSTHLSQSGETEVEETEETETVCHTHTYVITLPTQ---PS 756
Qy 758 PPAPV-----NKKTENVSCKLYLEKLYEFLNTSYICHKYILVSHSTMEKILKQ 806
Db 757 PPKEVYVVENSTEHSKNSDNOALTKVYLKLDLFTKSYCHKYILVSNSSMDQKLEEV 816
Qy 807 YKITEESKLSKSDPLDLNFIIONNIPVMSYSDLSLNSLSOLFMEIYKEMVCNLYK 866
Db 817 YNLTPKEENKLSKSDPLDLNFIIONNIPVMSYSDLSLNSLSOLFMEIYKEMVCNLYK 866
Qy 867 KDNKIKNLEBAKV-----STSVKTLSSS-----MQPLSLTTPDKPVS 908
Db 877 KEENHIKLLKEEQKQITSTSTSPGNTVTVAQSATHSNQONQSNASTNTONGVAVS 936
Qy 909 AND---DTSFS-----TNLNSKLFENILSLGKNKNIYQEL-IGKSSSENYEKILKSD 960
Db 937 SGPAVVEESHDPVLVLSISNDLKGIVSLNLGNKTKVNPPLTISTEMEKEVENILKND 996
Qy 961 TFYNESFTFVYKSRADDINLNDSEKRLLEEDINKLAKTLQLSFDLYNKYKLERLFD 1020
Db 997 TYFNDDIKQFYKSKSVITGUT-ETQKNALNDEIKKLKDTQLSFDLYNKYKLERLFD 1055
Qy 1021 KKKTVGYKMOIKKTLKLEQESKLSLNNPKHVLQNFVSFFNKKKEAEIAETENTLEN 1080
Db 1056 KKKELGQDKMOIKKTLKLEQESKLSLNNPKHVLQNFVSFFNKKKEAEIAETENTLEN 1115
Qy 1081 TKILLKHYKGLVYKNGESSPLKLTSEBSIOTEDNYASLENPKVLSKLEGLKDNLNLEK 1140
Db 1116 TKILLKHYKGLVYKNGESSPLKLTSEBSIOTEDNYANLEKFRVLSKIDGLNDNLHLGK 1175
Qy 1141 KKLVSLSGLHLHTAELKEVTKNNYTCNSPSENNTDNNNALESYKFLPEGTDVATVVS 1200
Db 1176 KKLVSLSGLHLHTAELKEVTKNNYTCNSPSENNTDNNNALESYKFLPEGTDVATVVS 1200
Qy 1201 E-----SGSDTLQEQOPKPPASTHYCAESNTITTSQNVDDVDVITVPIF 1246
Db 1235 PPQDVTSPISVRSVSGSGGSKTEETQIPTSGSLTELTQVQVQSONYDEEDDSLVVLPF 1294
Qy 1247 GESEEDYDGLQVYTGAVTPSVIDNLSKIENIEYEVLYLPLAGVYRSKQLKENNYMT 1306
Db 1295 GESEDNDEYDQVYTGAVT-NDNLSGFENEYDVYLYLPLAGVYRSKQLKENNYMT 1353
Qy 1307 FNVNWKDILNSRFNKNFNKVLSDILPYKDLTSSNVVYKDPYKFLNKEKRDKFLSSYN 1366
Db 1354 FNLNLDILNSRLKRRKFLVDSLDLMQFHHISSNEYIIEEDSFLLNSEQNTLLASYK 1413

Qy 1367 YIKSIDTDFINFANDVLGYKILSEKYSKSDLSIKKYYI-----NDK 1407
Db 1414 YIKSEVNDINFAGGEGISYKVLAKYKDDLESIKKVIKEKEFPSSPTTPPSAKTDE 1473
Qy 1408 QGENEKYLPFLNNTIETLYKTVNDKIDLEVIHLEAKVLYTEKSNVEKIKELNYLKTQ 1467
Db 1474 QKSEKFLPFTNTIETLYNVLNIDDLINLKAINDCNVEKDEAHVKITKLSOLKATD 1533
Qy 1468 DKLADFKNNFVGTADLSTDYNNHNLTKFLSTGWFENLAKTVLSNLGDLNLOGMLNI 1527
Db 1534 DKIDLFRNTDFEATKLIINDDTTKDMIGKLLSTGLV-ONFPNTIISKLEBGFODMLNI 1592
Qy 1528 SQHCYKVKQCPQNSGCFRHLDERECKCLLNYKQSGDKCVENPNPTCNENNGGCDADAKC 1587
Db 1593 SQHCYKVKQCPQNSGCFRHLDERECKCLLNYKQSGDKCVENPNPTCNENNGGCDADAKC 1652
Qy 1588 TEEDSGSNGKIKTCBCTKPDSPDPLDGFPCSSNFGISFLLILMLIYSFI 1639
Db 1653 TEEDSGSNGKIKTCBCTKPDSPDPLDGFPCSSNFGISFLLILMLIYSFI 1704

RESULT 2
Q9NHX1
ID Q9NHX1 PRELIMINARY; PRT; 1694 AA.
AC Q9NHX1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Major merozoite surface antigen.
GN GPI95.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC1/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and sequence analysis of major merozoite surface antigen(gp195)gene of plasmodium falciparum isolate FCC1/HN.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218248; AAF27526.1; -;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
SQ Merozoite.
KW SEQUENCE 1694 AA; 192767 MW; B51634A49E0F6728 CRC64;

Query Match 61.3%; Score 5162.5; DB 5; Length 1694;
Best Local Similarity 61.1%; Pred. No. 1.4e-153;
Matches 1054; Conservative 226; Mismatches 328; Indels 117; Gaps 27;

Qy 1 MKLIFFLCSLFFIINTOCVTHESYQELVKKLEALEDVLTGYSLFQEKMYLNEGTSGT 60
Db 1 MKLIFFLCSLFFIINTOCVTHESYQELVKKLEALEDVLTGYSLFQEKMYLNEGTSGT 60

Qy 61 AVTTSTPGSKGSVSGSGSVASGSGSVASGSGSVASGSGSGNSRRTPNSDSSD 120
Db 61 AVTTSTPGSKGSVSGSGSVASGSGSVASGSGSVASGSGSGNSRRTPNSDSSD 114

Qy 121 DAKSYADLKHVRVNYLLFTIKELYPQLFDLTNNHMLTCDNIHGFYKLYIDGYEIEINELLYK 180
Db 115 DAKSYADLKHVRVNYLLFTIKELYPQLFDLTNNHMLTCDNIHGFYKLYIDGYEIEINELLYK 174

Qy 181 LNFYDILLRAKINDVANDYCOIPENLKRANELDVLKLVFGYKPKLDNKNYKMGED 240
Db 175 LNFYDILLRAKINDVANDYCOIPENLKRANELDVLKLVFGYKPKLDNKNYKMGED 234

Qy 241 YIKNKKTIENINELIEESKKTIDKNKATKEEKKLYQAOYDLISYKQLEEAHNLI 300
Db 235 YIKNKKTIENINELIEESKKTIDKNKADNEEGKKLYQAOYDLISYKQLEEAHNLI 294

Qy 301 VLEKRIDTLKKNENIKELLDKINEIK-NPPANSANTPTNLLD-KNKKIEHEKEIK 358
Db 301 VLEKRIDTLKKNENIKELLDKINEIK-NPPANSANTPTNLLD-KNKKIEHEKEIK 358

```
Db 295 VLEKRIIDLTKKNNIKKLEIDIKITDAEKLTTGSKPNPLPENKKKEVEGHEEKIKEIA 354
Qy 359 KTIKKNIDSLFDPLEYLEYLRKKNKIDISAKVE--TKRESTEPN-EYPNGVTPYLSYND 415
Db 355 KTIKKNIDSLFDPLEYLEYLRKKNKVDVTPKSDQTPKVOIPKVPYNGVIVPLPLTD 414
Qy 416 INNAL---NELNSFGDLINPFYDTREPSKNIYTDN-ERKKFNEIKKIEKIKKKI----467
Db 415 IHNSLAADNKSNGDLMNP-DTKEKINEKIITDNKERKFINNIKKOIDLEKKNINHTK 473
Qy 468 ESKKSIEDRSKSLNDITKEYEKLLEIYDSKFNNDIDTNEFKEMKGRYSYKVKELTHH 527
Db 474 EQKKLEIDEDYERS---KKDYELLEKFYEMKFNNDKVDVDFKIFRSARYTYNVEKQKN 529
Qy 528 NTFASVENSKHLEKITKALKYMEDYSLRNIVVEKELKYKKNLISKIENIEITLVENIKK 587
Db 530 NFFSSNNSVYVQKLLKALSLEYDSLRKGISSEKDFNHYITLKTGLEADIKKLTTEEKS 589
Qy 588 DEEQLFEKK---ITKDKNKPDKEILEVSDIVKVOQVLLMNKIDELKKTQLILKNVELK 644
Db 590 SENKILEKNFKLTHSANAS---LEVSDIVKQVQVLLKKIEDLRKIELFLKNAULK 645
Qy 645 HNIHPNSYKQENKQEPYLLIVLKEIDKLKVPMPKVESLINEEKKNIKTEQSDNSEPS 704
Db 646 DSIHPNIIKQPNKPEPYLLIVLKEVDKLKEFIPKVDMLKKEQAVLSS-----I 696
Qy 705 TGEITGOATTPGOOAGSALGDSVQAOAQOQAPVPVPPEAKAQVTPPPAPV--762
Db 697 TQPLVAASSETDGGHSTHTLSQSGETEETETVGHITVITLPTQ---PSPKPEVKV 753
Qy 763 -----NNKTENVSKLDYLEKLYEFLNTSYICHKYLVSHTMNEKILKOYKITBEE 813
Db 754 VENSTEKSNDSNQAALTKTIVYLKDEFLTKSYICHKYLVSNSMDDQKLLVYNLTPEE 813
Qy 814 ESKLSCDPLDLLFNQNNIPWYSMFDSLNNLSOLFMEIYEKEMVNCVLYKLKONDKIK 873
Db 814 KNELKSCDPLDLLFNQNNIPAMYSLYDSMNNDLQHLFELYQKEMYYILHKLKEENHK 873
Qy 874 NLEEAQKV-----STSVKTLSSSS-----MQPLSLTPQDKPEVSAND---D 912
Db 874 KLLLEOKQITGTSSTSSPONTVNTAQSATHSNSNQSNASSTNTQNGVAVSSGPAVVE 933
Qy 913 TSHS-----TNLNSLAKLFENILSLGKNKIYQEL-IGQKSSNFENYKILKSDTTFYNESF 967
Db 934 ESHDPLTVLSISNDLKGIVSLLNLGKTKVPNPLATISTTEMEKFVENILKNNDTYFNDOI 993
Qy 968 TNFVSKADDINSLEKRRKLEEDINKLTKTLQSLPDLXNYKYLKLERLDFDKKTKVCK 1027
Db 994 KQFVNSKSVITGLT-ETQKNALNDEIKKLKDTLQSLFPLXNYKYLKLERLDFDKKTKVCK 1052
Qy 1028 YRMQIKKTLLEQLESKLSLNNPKHVLQNFVSFFNKKKEAEIAETENTLENTKILLKH 1087
Db 1053 DKMQIKKTLLEQLESKLSLNNPHNVLQNFVSFFNKKKEAEIAETENTLENTKILLKH 1112
Qy 1088 YKGLVKYNGESSPKLTLEESIQTDENTVASLENKVLKLEGLKLDNLEBKLLSYLS 1147
Db 1113 YKGLVKYNGESSPKLTLEESVSIQTDENTVANLEKFRVLSKIDGKLDNHLGKKLSFLS 1172
Qy 1148 SGLHLHIAELKEVKNKNTGSPSNNTDNNALLESYKKPELPGCTDVAIVYSE-----1201
Db 1173 SGLHLHIAELKEVKNKNTGSPSNNTDNNALLESYKKPELPGCTDVAIVYSE-----1201
Qy 1202 -----SGSDTLEQSQPKPASTHVGAESNTITTSQWDDVEDDVIIVPIFGESEEDY 1253
Db 1232 PPSLVRVSSSGSTKEETQITGSLTLEQVQSQNYDEEDSLVLPFIPGESEEDN 1291
Qy 1254 DDLGQVVTGEAVTPSVIDNLSKIENEYEVLYKPLAGVYRSLKQLENNVMTFNWVKD 1313
Db 1292 EYLDQVVTGEAVTSV-MDNILSGFENEYDVIYKPLAGVYRSLKQLENNVMTFNWVKD 1350
Qy 1314 ILSNRFENKBNKVNLESPLIPYKDLTSSNYYVKOPYFLNKEKRDCKFLSSNYIKDSTD 1373
Db 1351 ILSNRLKRRKRYFLDVLSEDLAQFKHSSNEYIIEDSFKLLNSEQNTLLKSYKIKESVE 1410
```

```
Qy 1374 TDINEANDVLGYKILSEKYKSDLSIKKYI-----NDKQGENEKY 1414
Db 1411 NDIKFAQEGISYEVKAKYKDDLESIKKVIKEEKEFPSSPTTPPSPAKTDQKKESKF 1470
Qy 1415 LPLFANNIETLYKTVNDKIDLFIHLEAKVLYNTYTKSNVEYKIKELNYLKTIODKLADPK 1474
Db 1471 LPLFNTIETLYNVLNKIDDIYLNLKAKIDNCNVEKDEAHVKITKLSDLKAIDDKIDLFPK 1530
Qy 1475 KNNFVGADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLDGNLQMLNLSOHCQVK 1534
Db 1531 NTNDFEATKLLNDOTKMDLKGKLLSTGLV-QNFPNTIISKLEIGKFDQMLNLSOHCQVK 1589
Qy 1535 KQCPNSGCFRHLDERECKCLLNKYQEGDKCVENPNTCNENNGCCDADAKTEEDSGS 1594
Db 1590 KQCPNSGCFRHLDERECKCLLNKYQEGDKCVENPNTCNENNGCCDADAKTEEDSGS 1649
Qy 1595 NGKKTCECTKPDSPVPLFDGIFCSCSNFGLIGISFLILMLILYSFI 1639
Db 1650 NGKKTCECTKPDSPVPLFDGIFCSCSNFGLIGISFLILMLILYSFI 1694

RESULT 3
Q9TZT5
ID Q9TZT5 PRELIMINARY; PRT; 1694 AA.
AC Q9TZT5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Merozoite surface protein 1.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HNI;
RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;
RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
RL Hainan, China."
RL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR EMBL; AF062348; AAC72884.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
SQ SEQUENCE 1694 AA; 192795 MW; 84CFC0E709F5673B CRC64;

Query Match 61.3%; Score 5162.5; DB 5; Length 1694;
Best Local Similarity 61.1%; Pred. No. 1.4e-153;
Matches 1054; Conservative 226; Mismatches 328; Indels 117; Gaps 27;
```

```
Qy 1 MKIIFLCFLFFIINTQCVTHESYQELVKLEALEDAVLTYSLFQKEKMLNEGTSCT 60
Db 1 MKIIFLCFLFFIINTQCVTHESYQELVKLEALEDAVLTYSLFQKEKMLNEGTSCT 60
Qy 61 AVTTSTPGSGSVASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTNPDSNDS 120
Db 61 AVTTSTPGSGSVASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTNPDSNDS 114
Qy 121 DAKSVADLKHRVNYLTITKELKYPOLFDLTNHLTLCDNIHGFYKYLIDGYEINELLYK 180
Db 115 DAKSVADLKHRVNYLTITKELKYPOLFDLTNHLTLCDNIHGFYKYLIDGYEINELLYK 174
Qy 181 LNFYDILLRAKLDNCANDYCOIPFNKLRANELDVLKLVFGYKRPPLDNKDNVKMED 240
Db 175 LNFYDILLRAKLDNCANDYCOIPFNKLRANELDVLKLVFGYKRPPLDNKDNVKMED 234
Qy 241 YTKKNNKTITENIELTEBSKTKIDKNKNATKBEKKKLYQAOYDLSIYNKQLEAHLIS 300
Db 235 YTKKNNKTITANIELTEBSKTKIDQKNADNBEKKKLYQAOYNLFIYNKQLEAHLIS 294
Qy 301 VLEKRIIDLTKKNNIKKLEIDIKITDAEKLTTGSKPNPLPENKKKEVEGHEEKIKEIA 358
Db 295 VLEKRIIDLTKKNNIKKLEIDIKITDAEKLTTGSKPNPLPENKKKEVEGHEEKIKEIA 354
```



```
QY 164 FYXLDGYBEINELYLKLFYFDLLRAKLVNDVANDYCOIPFNKLRANELDVLKLVFG 223
Db 179 FYXLDGYBEINELYLKLFYFDLLRAKLVNDVANDYCOIPFNKLRANELDVLKLVFG 238
QY 224 YRKPLDNIDNVGKMDYTKKNNKTITENINELIESKKTIDKNNKATKEEEKKLYQAOY 283
Db 239 YRKPLDNIDNVGKMDYTKKNNKTITIANINELIEGSKTIDQNKADNEEGKKLYQAOY 298
QY 284 DLSIYNKQLEEAHNLISVLEKRIDTLKKNENIKELDKINEIKNPPANSGTNPNTLLDK 343
Db 299 DLSIYNKQLEEAHNLISVLEKRIDTLKKNENIKELDKINEIKNPPANSGTNPNTLLDK 358
QY 344 NKKIEHEKEIEIAKTIFENIDSFTDPLELEYLREKKNIDISAKVE--TKESTEPN 401
Db 359 NKKIEHEKEIEIAKTIFENIDSFTDPLELEYLREKKNIDISAKVE--TKESTEPN 418
QY 402 EYPNGVTYPLSNDINNAL---NELNSFGDLINPDYTKEP-SKNIYTDN-ERKKFINE 455
Db 419 VYPNGIVYPLPLTDIHNSLAADNKNYSYGLMNP--HTKEKINEKLIITDNKERKIFINN 476
QY 456 IKKIDIEEKNHNTKEQNKLLDEYKS---KKDYBELLEKFEYEMKENNNFDRDQVVK 511
Db 477 IKKIDIEEKNHNTKEQNKLLDEYKS---KKDYBELLEKFEYEMKENNNFDRDQVVK 532
QY 512 MMGKRYSVYKVELTHNTFASVENSKHNLKLTALKYMEDYSLRNIVVYKELKYKNLI 571
Db 533 IFSARYTYNVEKQRYANKFSSNNYSVYQKLKALSLEYSLRKISGEKDFNHYITLK 592
QY 572 SKTENIEITLVENIKKDEOLEPEKK---ITKDNKPDKEILEVSDIVKVOVKVLLMKNKI 628
Db 593 TGLEADIKKLTKEIKSEENKILKPNKGLTHSANGS----LEVS DIVKVOVKVLLMKNKI 648
QY 629 DELKKTQLILKVNELKHNIHVPNSYKQENKQBPYYLIVLKKDEIDKLKVPMPKVESLINEE 688
Db 649 EDLRKIELFLKNAQLKDSIHVPNIYKPNQKPEYYLIVLKKDEIDKLKVPMPKVESLINEE 708
QY 689 KKNIKTEGOSDNSEPTGEITGOATTKPGCOQAGSALGDSVQAOAQEQKQAO-----P 742
Db 709 QAVLSS-----ITQPLVAASETTEDGGHSTHTLSOGETEVTVEETEETVGHHT 759
QY 743 PVPVPVPEAKAOVPTPPAPV-----NNKTENVSKLDYLEKLYEFLNYSYICHKY 791
Db 760 TVTITLPPQO---PSPPEKVKVYENSIEHKSNDNSQALTKYVLKKLDELFLKYSYICHKY 816
QY 792 ILVSHSTMKEKILKOYKITEESKLSKCDPLDLFNQNNIPVYMSFDSLNNLSQLF 851
Db 817 ILVSNSSMDQKLEVLNLTPEENELKSCDPLDLFNQNNIPVYMSFDSLNNLSQLF 876
QY 852 MEIYKEMVNCYLKLDNDKIKNLLEAKV-----STSVKTLSSSS----- 893
Db 877 FELYQEMTYLHLKKEENHKKLLEQKQITGTSSTSSPGNTVNTAQSAATHSNQOQ 936
QY 894 MOPLSLTPQDKPEVSAND---DTSHS---TNLNNSLKLFENILSLGKNKIYQEL- 945
Db 937 SNASSTNTQNGAVSGPVAWVEESHDPVLVLSINDLKGIVSLNNGKNTKVPNPJTIST 996
QY 946 KSENFENYKILKSDPTFYNESETFNFKSKADDINSLNDESRRKLEEDINKLTKTLQLSF 1005
Db 997 TEMEKYEYENILKNDYFNDDIKQFVKSNSKYITGLT-ETQKNALNDELKLDLTLQLSF 1055
QY 1006 DLYNKYKLERLFDKKTIVGYKMQIKKLTLLKEOLSKLSLNNPKHVLQNFVSFFNK 1065
Db 1056 DLYNKYKLERLFDKKTIVGYKMQIKKLTLLKEOLSKLSLNNPKHVLQNFVSFFNK 1115
QY 1066 KKEAEIAETENTLENPKILLKHYKGLVXYNGESSPLKTLSEESIQTESASLENFKVL 1125
Db 1116 KKEAEIAETENTLENPKILLKHYKGLVXYNGESSPLKTLSEESIQTESASLENFKVL 1175
QY 1126 SKLEGKLDNLNLEKKLSYLSGLHLIAELKEVTKNNKYTGNSPENNNTDVNNALESY 1185
Db 1176 SKIDGLNDNLHLGKKLSYLSGLHLIAELKEVTKNNKYTGNSPENNNTDVNNALESY 1235
QY 1186 KFLPEGTGTOAVTVVSE-----SGSDTLQEQSQPKKPASTHVGAESNTITTSQ 1231
```

```
Db 1236 ENFLPE-AKVTTVTPPQPDVTPSPLSVRVSGSGSTKEETQIPTSGSLTLLQVVQLO 1294
QY 1232 NYDDEVDVVIIPIFGESEEDYDGLGVVTPGEAVTPSVIDNLSKIENIEVYLYKPLAG 1291
Db 1295 NYDEEDSVLVIPIFGESEEDNDEYLDQVVVTEGASVIT-MDNILSGFENYDYVLYKPLAG 1353
QY 1292 VYRSKQLQLENNVMTFNVNVDILNSRKNFNKVLNEDSLIPYKDLTSSNYYVVDYK 1351
Db 1354 VYRSKQLQLENNVMTFNVNVDILNSRKNFNKVLNEDSLIPYKDLTSSNYYVVDYK 1413
QY 1352 FLNKKRDKFLSYNYIKDSITDINFANDVLYGYKILSEKYSKLDLSKKYI----- 1404
Db 1414 LLNSEQNTLLSKYIKESVENDIKFAQEGISYSEKVLAKYKDDLESIKKVIKEKEKF 1473
QY 1405 -----NDKQGENEKYLPFLNNIETLYKTVDNKIDLVHLHLEAKVLNLYEKS 1451
Db 1474 PSSPTTTPSPAKTDQKKESKFLPLTWIETLYNNLVNKKIDYLLNLAKKINDCNVEKD 1533
QY 1452 NYEVIKELNYLKTODKLADEFKNNNFVGIADLSTDYNNHNLTKPLSTGMVFENLAKT 1511
Db 1534 EAHVITKLSDLKAIDDKIDLFKNPYDFAIKKLINDTKDMLGKLLSTGLV-QNFPNT 1592
QY 1512 VLSNLDGNLQGNLNTISOHQYKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPN 1571
Db 1593 IISKIEGKFDQMLNTISOHQYKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPN 1652
QY 1572 PTCNENGGCCADAKTEEDSGNSGKKITCECTKPDSPYPLFDGIFGSSNFIPLLL 1631
Db 1653 PTCNENGGCCADAKTEEDSGNSGKKITCECTKPDSPYPLFDGIFGSSNFIPLLL 1712
QY 1632 MLILYSFI 1639
Db 1713 MLILYSFI 1720

RESULT 5
Q8T6A9 PRELIMINARY; PRT; 656 AA.
ID Q8T6A9 AC Q8T6A9,
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Merozoite surface protein 1 (fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCB-1;
RA Li X.R., Liu S.C., Chishti A.H., Oh S.S.;
RT "83 kDa subfragment of MSP-1 gene of Plasmodium falciparum isolate FCB-1."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF480451; AAL8455.1;
FT NON_TER 1
FT TER 656
SQ SEQUENCE 656 AA; 75698 MW; ELEF44ACE341FD0C9 CRC64;

Query Match 39.9%; Score 3361; DB 5; Length 656;
Best Local Similarity 100.0%; Pred. No. 7.6e-98;
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 VTHESYQELVKKLEALEDVAVLTGYSLFQEKVYLVNLSGTSTAVTTSTPGSKGSVASGGSG 79
Db 1 VTHESYQELVKKLEALEDVAVLTGYSLFQEKVYLVNLSGTSTAVTTSTPGSKGSVASGGSG 60
QY 80 GSVASGGSVASGGSVASGGSVASGGSNSRRTPNSDSDSDAKSYADLKHRYNRYLLTI 139
Db 61 GSVASGGSVASGGSVASGGSVASGGSNSRRTPNSDSDSDAKSYADLKHRYNRYLLTI 120
QY 140 KELKYPQLFDNLNMLTCLDNIHGFKYLIDGVEENELLYKLNIFYDILLRAKLVNDV 199
```



```
Db 121 KELKYPQLFDLTHMLTLCNDTHGFKYLDIGYEINELLYKLFYDGLRAKLDVNCAND 180
QY 200 YCOIPFNLIKIRANEDVLKLVFGYRKPLDNLKDNVGMEDYIKKNNKTIENINELIES 259
Db 181 YCOIPFNLIKIRANEDVLKLVFGYRKPLDNLKDNVGMEDYIKKNNKTIENINELIES 240
QY 260 KKTIDKNKATKEEKKKLYQAYDLSYINKOLEEAHNLISVLEKRIDTFLKKNENIKELL 319
Db 241 KKTIDKNKATKEEKKKLYQAYDLSYINKOLEEAHNLISVLEKRIDTFLKKNENIKELL 300
QY 320 DKINEIKNPPANGNTPTNLTKNNKKEEHEKEKEIAKTIKFNIDSFTDPLEEYLYL 379
Db 301 DKINEIKNPPANGNTPTNLTKNNKKEEHEKEKEIAKTIKFNIDSFTDPLEEYLYL 360
QY 380 REKNKIDISAKVETKESTPEPNEYNGVTPYPLSYNDINNALNELNSFGDLINFPDYTKEP 439
Db 361 REKNKIDISAKVETKESTPEPNEYNGVTPYPLSYNDINNALNELNSFGDLINFPDYTKEP 420
QY 440 SKNIYTDNERKKFINEIKIEKIKIESDKKSYEDRSKSLNDITKEYEKLLEIYDSK 499
Db 421 SKNIYTDNERKKFINEIKIEKIKIESDKKSYEDRSKSLNDITKEYEKLLEIYDSK 480
QY 500 FNNIDLTNFKMMGKRYSYKVEKLTHTNTPFASYENSKHNLEKLTALKYMEDYSLRNIV 559
Db 481 FNNIDLTNFKMMGKRYSYKVEKLTHTNTPFASYENSKHNLEKLTALKYMEDYSLRNIV 540
QY 560 VEKELKYKNLISKIENETIETLVENIKKDEOLFEEKITKDNKPKDEKILEVSDIVKQV 619
Db 541 VEKELKYKNLISKIENETIETLVENIKKDEOLFEEKITKDNKPKDEKILEVSDIVKQV 600
QY 620 QKVLLMKNIDELKKTQLLILKNVELKHNHVPNSYKQENKQEPYIYLVKKKEIDKILK 675
Db 601 QKVLLMKNIDELKKTQLLILKNVELKHNHVPNSYKQENKQEPYIYLVKKKEIDKILK 656

RESULT 6
Q26194 ID Q26194 PRELIMINARY: PRT: 1751 AA.
AC Q26194;
DC 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DI 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE P200 protein precursor.
GN P200.
OS Plasmodium vivax.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5855;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SAL-1;
RX MEDLINE=92158013; PubMed=1371329;
RA Gibson H.L., Tucker J.E., Kaslow D.C., Krettl A., Collins W.E.,
Kiefer M.C., Bathurst I.C., Barr P.J.;
RT "Structure and expression of the gene for P200, a major blood-stage
surface antigen of plasmodium vivax."
RL Mol. Biochem. Parasitol. 50:325-334(1992).
DR EMBL; M75674; AAA29735.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Signal.
FT SIGNAL.
FT CHAIN
SQ SEQUENCE 1751 AA; 196120 MW; 4162F9883F29D8A6 CRC64;
```

Query Match 36.7%; Score 3093.5; DB 5; Length 1751;
Best Local Similarity 38.2%; Pred. No. 4.4e-89;
Matches 704; Conservative 316; Mismatches 517; Indels 307; Gaps 41;

```
QY 1 MKIIFLCSFLPIINTOCVTHESYQELVKLEALEDAVLTCYSLFOKEKMYLNEGTSCT 60
Db 1 MKALLFLSFIFVTKOCET-ESTKQLVAKLDKLEALVVDGTYELFHKKL----- 50
```

```
QY 61 AVTTSTPGSKGVSAGSGSGSVASGSGSVASGSGSVASGSGSGSRRTNPSONSDS 120
Db 51 -----GENDIKVETNASANNNN 68
QY 121 DAKSVADUKHRVRYNLLTTIKELKYPQLFDLTHMLTLCNDTHGFKYLDIGYEINELLYK 180
Db 69 NOVSV--LTSKRNFLSKFLEQLPGHTOLLHLIRELAVEPNGIKYLVSEYEFNQLMHV 126
QY 181 LNFYDGLRAKLDVNCANDYCOIPFNLIKIRANEDVLKLVFGYRKPLDNLKDNVGMED 240
Db 127 INFHDDLRAKLDVNCANDYCOIPFNLIKIRANEDVLKLVFGYRKPLDNLKDNVGMED 186
QY 241 YKKKKTKTIENEL-IEESKK----- 261
Db 187 FITKKKIITIKNISDLIIAENKKRSGHPTTTNAGCTQPANGSTAAASSETTQISGSSNSG 246
QY 262 -----TIDKNNKATKEEKKKLYQAYDLSYINKQ 291
Db 247 SSSTGSSSGSSSTGSGTGTGQSPPAADASSTNANYEAKTIYQAVNTIFTYQ 306
QY 292 LEEAHNLISVLEKRIDTFLKKNENIKELDKINEIKNPPA---NSGNTPTNLTKNNKKE 348
Db 307 IQEAKLTLAVLEKRVKVLKEHKDIKVLLEQVAKENEKLPDYPNTNLTNVHKEAESKIA 366
QY 349 EHEKEIKETAKTIKFNIDSFTDPLEEYLYREKKNKI-DISAKVETKESTPEPNE----- 402
Db 367 ELEKKIEAKIATVDFDGLFTDAELEYLYREKAKMAGTLIIPESTKSAGTGVPTL 426
QY 403 ---YNGVTPYPLSYNDINNALNELNS---FGDLINFPDYTKEPSKI-YTDNERKKFINE 455
Db 427 KETYPHGYSYLAENSIVLEIKIGSDETFGLQNP-DDGKQPKKGLILNETKRKELLEK 485
QY 456 IKEKIKIEKKTESDKSYEDRSKSLNDITKEYEKLLEIYDSKFNNDLTNFKEMCK 515
Db 486 IMNKIKIEDKLPNLKKEEYKLYEAKVNEFKPAPNFHYEARLONTLVENKFDDEK 545
QY 516 RYSYKVEKLTHTNTPFASYENSKHNLEKLTALKYMEDYSLRNIVVEKELKYKNLISKIE 575
Db 546 REAYMEKKKLES--CSYEQNSNLINKKQLTLEDYVLRKDIADDEIKHESFMEMK 603
QY 576 NIEITLVENIKKDEOLFEEKITKDNKPKDEKILEVSDIVKQVQVLLMKNIDELKKTQ 635
Db 604 SEIYDLAQEIRKNNKL-----TIENK-----FDFSGVVELOVQVLIKIEALKNYQ 652
QY 636 LILKNVELKHNHVPNSYKQENKQEPYIYLVKKKEIDKILKLVEMPKVESLINEBEK 690
Db 653 NLLKNAKVKDDLYIPKYKTSKPEPYIYLVKKREIDKLDKDFIPKIESMIATEKKNPTVA 712
QY 691 ---NIKTEGOS--DNSEPSTEG-----EITQATTKPG-----QAQSA-LE 726
Db 713 AADIVAKGOSLRGASSETGTTGNTVNAQTAVQVQHVNVNAVTVQPGTTGHOAGGEAETQ 772
QY 727 GDSVQAQAOEQKQA-----OPVPVPV-PE-AKAQVTPPAPVNNKTNVSK 771
Db 773 TNSVQAQVQQTTPAGAGGOVASTQISOAPATQASPEPAAPSTPAAAAPAPTMSK 832
QY 772 LDYLEKLYEFLNTSYCHKYILVSHSTNMEKILKQYKITEEESKL--SSCDPLDLFFNI 829
Db 833 LEYLEKLLDFLAKSAYACHKHIFVTSNMTKKELLDQYKLNADQONKINETKCDLDELFFNV 892
QY 830 QNNIPVMYSMFDLSNLSQLFWEIYKEMVGNLYKLDND-KIKNLLSEAK-----VS 883
Db 893 QNNLPAMYIYDSMSNLQNLIELYQKEMVNIYKNDTKKIRAFLETLGSKAAAPAQ 952
QY 884 TSVKTLSSSSMOPLSIT-PQDKPEVSANDTSHSTNLNLSKLFIENILSLGKNKIYQEL 942
Db 953 SAAKPSGQAGTTPVTPTAPVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1010
QY 943 IGKKSSENFYEKILKDSFTFYNESTFNFKSKADDTINSLNDESRRKKLEEDINKLKTITQ 1002
Db 1011 ADTAQVEFYKHLSQLDK-YNDYQKFLSOKDEITKM-DETKWALCAEIEELKKLQ 1068
QY 1003 LSFIDLNYKYLKLERLFDKKKTGVGYKMOIKKLTLLEQLESKLSLNNPKHVLFQNSVF 1062
```


Db	1457	GELYKTHLAVNEEVKKVEADIKAEDDKIKKIGSDSTKTEKTQSMKRAAELEKYLPPFLN	1516			
Qy	1420	NIELTKYTVNDKIDLFVIHLEAKVLYTYEKSNEVYKIKELNLYKTIQDKLADFKNNPF	1479			
Db	1517	SLOKEYESLVSKVNTYTDNLKVVINNCQLEKKEAEITVKKLQDYNKMDKLEBYKK----	1572			
Qy	1480	VGIADLSTQYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQGMNLISOHOCVKKOCPO	1539			
Db	1573	---SEKNEVKSSGLEKMLKSLIKENESKEILSOLLNVQTLTMSSEHCIDTNVDP	1629			
Qy	1540	NSGCFRHLDERECKLLNYKQEGDKCVENPNTCENNNGGCDADAKCTEEDSGNGKKI	1599			
Db	1630	NAACYRYLDMEBWRCLLTFKEBGGKCVPGSNVTCCKDNGGCAPEACECKWTD-----	1685			
Qy	1600	TCBCTKPDSPYPLFDGIFCSCSSNFGISFLILMLIL	1635			
Db	1686	VCKCTKGESEPLFEGVFCSSSFLSLFLMLLFL	1721			
RESULT 9						
Qy	Q9TYG1	PRELIMINARY;	PRT;	539	AA.	
AC	Q9TYG1					
DT	01-MAY-2000 (Tremblrel. 13, Created)					
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)					
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)					
DE	Major merozoite surface protein (Fragment).					
GN	MSPI.					
OS	Plasmodium falciparum.					
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
OX	NCBI_TaxID=5833;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=93295445; PubMed=8515786;					
RA	Jongwutives S., Tanabe K., Kanbara H.;					
RT	"Sequence conservation in the C-terminal part of the precursor to the					
RT	major merozoite surface proteins (MSPI) of Plasmodium falciparum from					
RT	field isolates."					
RL	Mol. Biochem. Parasitol. 59:95-100(1993).					
DR	EMBL; D13358; BAA02619.1; -.					
DR	InterPro; IPR000561; EGF-like.					
DR	Pfam; PF00008; EGF; 1.					
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.					
KW	EGF-like domain; Merozoite.					
FT	NON_TER					
SQ	SEQUENCE 539 AA; 61075 MW; C70C2E100EC4A101 CRC64;					
Query Match						
Best Local Similarity 33.3%; Score 2801; DB 5; Length 539;						
Matches 539; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
Qy	1101	PLKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKKLKLSYSSGLHHLIAELKEV	1160			
Db	1	PLKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKKLKLSYSSGLHHLIAELKEV	60			
Qy	1161	IKNKNTGNSPSNNTDVNNALLESYKKFLPEGTDVATVYSESGSDTLEOSQPKKPASTHV	1220			
Db	61	IKNKNTGNSPSNNTDVNNALLESYKKFLPEGTDVATVYSESGSDTLEOSQPKKPASTHV	120			
Qy	1221	GAESNTITTSQNVDDDEVDDVVIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE	1280			
Db	121	GAESNTITTSQNVDDDEVDDVVIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE	180			
Qy	1281	YEVLYLKLPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKNFNKLVLESLLPYKDLT	1340			
Db	181	YEVLYLKLPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKNFNKLVLESLLPYKDLT	240			
Qy	1341	SSNVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYGKILSEKYSKLDLSI	1400			
Db	241	SSNVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYGKILSEKYSKLDLSI	300			
Qy	1401	KYINDKQGENEKYLPFLANNIETLYKTVDKIDLFVIHLEAKVLYTYEKSNEVYKIKEL	1460			
Db	301	KYINDKQGENEKYLPFLANNIETLYKTVDKIDLFVIHLEAKVLYTYEKSNEVYKIKEL	360			

Qy	1401	KYINDKQGENEKYLPFLANNIETLYKTVDKIDLFVIHLEAKVLYTYEKSNEVYKIKEL	1460			
Db	301	KYINDKQGENEKYLPFLANNIETLYKTVDKIDLFVIHLEAKVLYTYEKSNEVYKIKEL	360			
Qy	1461	NYLKTIQDKLADFKKNNFVGITADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGN	1520			
Db	361	NYLKTIQDKLADFKKNNFVGITADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGN	420			
Qy	1521	LOGMLNISOHQCVKQKOPNSGCFRHLDERECKLLNYKQEGDKCVENPNTCENNNGG	1580			
Db	421	LOGMLNISOHQCVKQKOPNSGCFRHLDERECKLLNYKQEGDKCVENPNTCENNNGG	480			
Qy	1581	CDADAKCTBEDSGNGKKITCECTKPDSPYPLFDGIFCSCSSNFGISFLILMLILYSFI	1639			
Db	481	CDADAKCTBEDSGNGKKITCECTKPDSPYPLFDGIFCSCSSNFGISFLILMLILYSFI	539			
RESULT 10						
Qy	Q25972	PRELIMINARY;	PRT;	539	AA.	
AC	Q25972;					
DT	01-NOV-1996 (Tremblrel. 01, Created)					
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)					
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)					
DE	Major merozoite surface protein (Fragment).					
GN	MSPI.					
OS	Plasmodium falciparum.					
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
OX	NCBI_TaxID=5833;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=93295445; PubMed=8515786;					
RA	Jongwutives S., Tanabe K., Kanbara H.;					
RT	"Sequence conservation in the C-terminal part of the precursor to the					
RT	major merozoite surface proteins (MSPI) of Plasmodium falciparum from					
RT	field isolates."					
RL	Mol. Biochem. Parasitol. 59:95-100(1993).					
DR	EMBL; D13360; BAA02621.1; -.					
DR	InterPro; IPR000561; EGF-like.					
DR	InterPro; IPR001245; Tyr_pkinase.					
DR	Pfam; PF00008; EGF; 1.					
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.					
KW	EGF-like domain; Merozoite.					
FT	NON_TER					
SQ	SEQUENCE 539 AA; 61076 MW; 72186F1412C57D1C CRC64;					
Query Match						
Best Local Similarity 33.2%; Score 2798; DB 5; Length 539;						
Matches 538; Conservative 1; Mismatches 0; Indels 0; Gaps 0;						
Qy	1101	PLKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKKLKLSYSSGLHHLIAELKEV	1160			
Db	1	PLKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKKLKLSYSSGLHHLIAELKEV	60			
Qy	1161	IKNKNTGNSPSNNTDVNNALLESYKKFLPEGTDVATVYSESGSDTLEOSQPKKPASTHV	1220			
Db	61	IKNKNTGNSPSNNTDVNNALLESYKKFLPEGTDVATVYSESGSDTLEOSQPKKPASTHV	120			
Qy	1221	GAESNTITTSQNVDDDEVDDVVIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE	1280			
Db	121	GAESNTITTSQNVDDDEVDDVVIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE	180			
Qy	1281	YEVLYLKLPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKNFNKLVLESLLPYKDLT	1340			
Db	181	YEVLYLKLPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKNFNKLVLESLLPYKDLT	240			
Qy	1341	SSNVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYGKILSEKYSKLDLSI	1400			
Db	241	SSNVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYGKILSEKYSKLDLSI	300			
Qy	1401	KYINDKQGENEKYLPFLANNIETLYKTVDKIDLFVIHLEAKVLYTYEKSNEVYKIKEL	1460			
Db	301	KYINDKQGENEKYLPFLANNIETLYKTVDKIDLFVIHLEAKVLYTYEKSNEVYKIKEL	360			


```
|||||
Db 421  LOGMLNISOHQVKKQCPQNSGCGFRHLDERECKLLNFKQSGKCVENPNTCENNGG 480
|||||
QY 1581  CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLLLILMLILYSFI 1639
|||||
Db 481  CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLLLILMLILYSFI 539
|||||

RESULT 13
Q25976 PRELIMINARY; PRT; 539 AA.
AC Q25976;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13362; BAA02623.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61045 MW; 8BF9C64322E9A778 CRC64;

Query Match 33.2%; Score 2794; DB 5; Length 539;
Best Local Similarity 99.8%; Pred. No. 3e-80;
Matches 538; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKTLSESIQTEDNVASLENFKVLSKLEGLKDNLNLEKKKLSYSSGLHLLIAELKEV 1160
Db 1 PLKTLSESIQTEDNVASLENFKVLSKLEGLKDNLNLEKKKLSYSSGLHLLIAELKEV 60
QY 1161 IKKNKNTGNSPSNNTDVNNALSYKKFLPEGTDTVATVVSSESGDITLQSQPKKPASTHV 1220
Db 61 IKKNKNTGNSPSNNTDVNNALSYKKFLPEGTDTVATVVSSESGDITLQSQPKKPASTHV 120
QY 1221 GAESENTITTSQNVDDVDVVIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE 1280
Db 121 GAESENTITTSQNVDDVDVVIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE 180
QY 1281 YEVLKPLAGVYRSKQLKLENNVMTFNVNVDILNSRFNKNFNKRVNLESILPYKDLT 1340
Db 181 YEVLKPLAGVYRSKQLKLENNVMTFNVNVDILNSRFNKNFNKRVNLESILPYKDLT 240
QY 1341 SSVVVKDPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGYKILSEKYSKSLDSI 1400
Db 1341 SSVVVKDPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGYKILSEKYSKSLDSI 300
QY 1401 KKYINDKQGENEYLPFLNNIETLYKTVNDKIDLFIHLEAKVNLNTYKSNVEVKIKEL 1460
Db 301 KKYINDKQGENEYLPFLNNIETLYKTVNDKIDLFIHLEAKVNLNTYKSNVEVKIKEL 360
QY 1461 NYLKTIDQKLADFKKNNFVGIALSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 1520
Db 361 NYLKTIDQKLADFKKNNFVGIALSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 1521 LQGLMNLISQHQVKKQCPQNSGCGFRHLDERECKLLNFKQSGKCVENPNTCENNGG 1580
Db 421 LQGLMNLISQHQVKKQCPQNSGCGFRHLDERECKLLNFKQSGKCVENPNTCENNGG 480
|||||
QY 1581  CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLLLILMLILYSFI 1639
|||||
Db 481  CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLLLILMLILYSFI 539
|||||
```

```
QY 1581  CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLLLILMLILYSFI 1639
|||||
Db 481  CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLLLILMLILYSFI 539
|||||

RESULT 14
Q25981 PRELIMINARY; PRT; 539 AA.
AC Q25981;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13363; BAA02624.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61046 MW; 3ED87473EE87B65 CRC64;

Query Match 33.1%; Score 2791; DB 5; Length 539;
Best Local Similarity 99.6%; Pred. No. 3.7e-80;
Matches 537; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKTLSESIQTEDNVASLENFKVLSKLEGLKDNLNLEKKKLSYSSGLHLLIAELKEV 1160
Db 1 PLKTLSESIQTEDNVASLENFKVLSKLEGLKDNLNLEKKKLSYSSGLHLLIAELKEV 60
QY 1161 IKKNKNTGNSPSNNTDVNNALSYKKFLPEGTDTVATVVSSESGDITLQSQPKKPASTHV 1220
Db 61 IKKNKNTGNSPSNNTDVNNALSYKKFLPEGTDTVATVVSSESGDITLQSQPKKPASTHV 120
QY 1221 GAESENTITTSQNVDDVDVVIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE 1280
Db 121 GAESENTITTSQNVDDVDVVIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE 180
QY 1281 YEVLKPLAGVYRSKQLKLENNVMTFNVNVDILNSRFNKNFNKRVNLESILPYKDLT 1340
Db 181 YEVLKPLAGVYRSKQLKLENNVMTFNVNVDILNSRFNKNFNKRVNLESILPYKDLT 240
QY 1341 SSVVVKDPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGYKILSEKYSKSLDSI 1400
Db 241 SSVVVKDPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGYKILSEKYSKSLDSI 300
QY 1401 KKYINDKQGENEYLPFLNNIETLYKTVNDKIDLFIHLEAKVNLNTYKSNVEVKIKEL 1460
Db 301 KKYINDKQGENEYLPFLNNIETLYKTVNDKIDLFIHLEAKVNLNTYKSNVEVKIKEL 360
QY 1461 NYLKTIDQKLADFKKNNFVGIALSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 1520
Db 361 NYLKTIDQKLADFKKNNFVGIALSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 1521 LQGLMNLISQHQVKKQCPQNSGCGFRHLDERECKLLNFKQSGKCVENPNTCENNGG 1580
Db 421 LQGLMNLISQHQVKKQCPQNSGCGFRHLDERECKLLNFKQSGKCVENPNTCENNGG 480
|||||
QY 1581  CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLLLILMLILYSFI 1639
|||||
Db 481  CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLLLILMLILYSFI 539
|||||
```

```
RESULT 15
Q25984
ID Q25984 PRELIMINARY; PRT; 539 AA;
AC Q25984;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13356; BAA02617.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61115 MW; 3788015F3127CB9E CRC64;

Query Match 33.1%; Score 2789; DB 5; Length 539;
Best Local Similarity 99.6%; Pred. No. 4.2e-80;
Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIQTEDNYASLENFKVLSKLEGKLDNLEKKKLSYLSGSLHLLIAELKEV 1160
Db 1 PLKLTSEESIQTEDNYASLENFKVLSKLEGKLDNLEKKKLSYLSGSLHLLIAELKEV 60

QY 1161 INKNYTGNSPSNNTDVNNALESYKFLPEGTDTATVVSSESDTLEQSQPKKPASTHV 1220
Db 61 INKNYTGNSPSNNTDVNNALESYKFLPEGTDTATVVSSESDTLEQSQPKKPASTHV 120

QY 1221 GAESNTITTSQNVDDVDVVIIVPIFGESEEDYDDLQGVVTEAVTPSVIDNLSKIENE 1280
Db 121 GAESNTITTSQNVDDVDVVIIVPIFGESEEDYDDLQGVVTEAVTPSVIDNLSKIENE 180

QY 1281 YEVLYLKLPLAGYVRSLLKOLENNVMTFNVNVDILNSRFNKNVLESDLIPYKDLT 1340
Db 181 YEVLYLKLPLAGYVRSLLKOLENNVMTFNVNVDILNSRFNKNVLESDLIPYKDLT 240

QY 1341 SSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYKILSEKYSKSDLSI 1400
Db 241 SSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYKILSEKYSKSDLSI 300

QY 1401 KKYINDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLYNTEKSNVEVKIKEL 1460
Db 301 KKYINDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLYNTEKSNVEVKIKEL 360

QY 1461 NYLKTITQDKLADPKNNNFVGLADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 1520
Db 361 NYLKTITQDKLADPKNNNFVGLADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 420

QY 1521 LOGMLNISQHCVKQKQPCQNSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGG 1580
Db 421 LOGMLNISQHCVKQKQPCQNSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGG 480

QY 1581 CDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCSSNFGISFLILMLILYSFI 1639
Db 481 CDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCSSNFGISFLILMLILYSFI 539

RESULT 16
Q25645
ID Q25645 PRELIMINARY; PRT; 1787 AA.
AC Q25645; 000883;
```

```
Q25971
ID Q25971 PRELIMINARY; PRT; 539 AA.
AC Q25971;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13359; BAA02620.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61017 MW; 5CA9C651BB62B5E1 CRC64;

Query Match 33.1%; Score 2788; DB 5; Length 539;
Best Local Similarity 99.6%; Pred. No. 4.6e-80;
Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIQTEDNYASLENFKVLSKLEGKLDNLEKKKLSYLSGSLHLLIAELKEV 1160
Db 1 PLKLTSEESIQTEDNYASLENFKVLSKLEGKLDNLEKKKLSYLSGSLHLLIAELKEV 60

QY 1161 INKNYTGNSPSNNTDVNNALESYKFLPEGTDTATVVSSESDTLEQSQPKKPASTHV 1220
Db 61 INKNYTGNSPSNNTDVNNALESYKFLPEGTDTATVVSSESDTLEQSQPKKPASTHV 120

QY 1221 GAESNTITTSQNVDDVDVVIIVPIFGESEEDYDDLQGVVTEAVTPSVIDNLSKIENE 1280
Db 121 GAESNTITTSQNVDDVDVVIIVPIFGESEEDYDDLQGVVTEAVTPSVIDNLSKIENE 180

QY 1281 YEVLYLKLPLAGYVRSLLKOLENNVMTFNVNVDILNSRFNKNVLESDLIPYKDLT 1340
Db 181 YEVLYLKLPLAGYVRSLLKOLENNVMTFNVNVDILNSRFNKNVLESDLIPYKDLT 240

QY 1341 SSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYKILSEKYSKSDLSI 1400
Db 241 SSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYKILSEKYSKSDLSI 300

QY 1401 KKYINDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLYNTEKSNVEVKIKEL 1460
Db 301 KKYINDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLYNTEKSNVEVKIKEL 360

QY 1461 NYLKTITQDKLADPKNNNFVGLADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 1520
Db 361 NYLKTITQDKLADPKNNNFVGLADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 420

QY 1521 LOGMLNISQHCVKQKQPCQNSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGG 1580
Db 421 LOGMLNISQHCVKQKQPCQNSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGG 480

QY 1581 CDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCSSNFGISFLILMLILYSFI 1639
Db 481 CDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCSSNFGISFLILMLILYSFI 539

RESULT 17
Q25645
ID Q25645 PRELIMINARY; PRT; 1787 AA.
AC Q25645; 000883;
```



```
QY 1428 VNDKIDLFVIHLEAKVNTYKESNVKIKELNYLKTIOQKLADPKKN--NNFVGIADLUS 1486
Db 1581 ILQSEEEYIEVLQNRDLSYKNEKTEFEILTUKNLEYIKIDEKLENFVENAENKHIASTA 1640
QY 1487 TDYNNHNLTKFLSGWPFENLAKTVLSNLDGNLQGLMNLIS-OHQCVK-KOCPONSGCF 1544
Db 1641 L-----NNL-----NKSGLVGESESKILAKMLNMDSDLLGIDPKHVCINTRDIPANAGCF 1692
QY 1545 RHLDREBECKLLNTYKQEGDKCVENPNPTCNENNGCGDADAKCTEDSDSGSKKKTICBCT 1604
Db 1693 RYDNGNEERWCLGKYKNNNTCIEDSNPTCGNNGGCDPTAGCQTAENRENSKKIICTCK 1752
QY 1605 KPDSYPLFGICFSSNPLGIGLFLILLMLILYS 1637
Db 1753 EPTPNAYDGVFCSSSPGSLIILLITLIVEN 1785

RESULT 18
Q9U6D4
ID Q9U6D4 PRELIMINARY; PRT; 1791 AA.
AC Q9U6D4;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface antigen 1.
OS Plasmodium berghei (strain Anka).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANKA;
RA Zhong H., Fan J.Y., Yang S., Davidson E.A.;
RT "Cloning and characterization of the merozoite surface antigen 1 gene of Plasmodium berghei.";
RL Am. J. Trop. Med. Hyg. 60:994-999(1999).
DR EMBL; AF187232; AAF13063.1; -.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICEXTENSN.
SQ SEQUENCE 1791 AA; 198568 MW; 0474251901CC6A99 CRC64;

Query Match 30.0%; Score 2530; DB 5; Length 1791;
Best Local Similarity 31.8%; Pred. No. 1.7e-71;
Matches 604; Conservative 339; Mismatches 586; Indels 368; Gaps 40;

QY 1 MKIIIFLCSFLFIINTQCVTHESQBELVKLEALEDAVLTCYSLFQEKVMNLECTST 60
Db 1 MKVIGLLFSVFFAICKSETIEVNDIIIOKLEKLESUVEGLELUFQKSVQIIN----- 54
QY 61 AVTTSTPGSKGVSAGSGGVSAGSGSVASGSGSVASGSGGSGNSRRTPN-SDNSSD 119
Db 55 -----ASPPSETINPESDNT--- 69
QY 120 SDAKSYADUKHRVNVLLTIKELKYPQLPDLNHHMLTLCNTHGPKYLDGYEENELLY 179
Db 70 -----PAPKLGQFITFEELGETEQLVNLTKTLVFNKYGLKYLIESKEEFNGLMH 121
QY 180 KLNFFDLRLAKINDVCANDYQIIPENLAKIRANELDVLKLVFGYRKPDLNDKDNVGM 239
Db 122 ANFYDVRDLNDMCANCYCEIPEHLKISBEETEMLKKVILGVYRKPENIQQDIEKLE 181
QY 240 DYIKNNKKTNIENELI-BESKRTIDKNK-----NATKEEEKKKYQAQVDLSIYNK 290
Db 182 IYITKNKEIVTALNTLIAETKKITPENADCNDNCTDESKYSKKKIIYQAMYNVIFYK 241
QY 291 QLEEAHNLVSLEKRIDTLTKKNENIKELDDKINEITKNPP-----PANGS 334
Db 242 QLAIEQKVIETLEKRYSAKKNDVINKPLLOQIEDIAKAAVPTTEGQITTSQSQSSTEP 301
QY 335 NTPN----- 338
Db 302 -TPSSGEVSTGTSGASAGVTNTGAATGTGTGAATGTGTGAATGTGTGAATGTGCAEAATG 360
```

```
QY 339 -----TLLDNKKIEBEKEIKEIAKTIKFNIDSLFTDPLELEYLYREKN 383
Db 361 NNTNTEVTVQVPTLTPEKKKKMDGLYQAQIKIEIAKTIKFNLDGIFVNFIELEYFKKEK 420
QY 384 K---NIDISA--KVETKESTEP--NEYPNGVTYPLSYNDI-----NNALNELNSFGDLIN 431
Db 421 KESCNLSTSSCKKNKTSETIIPLVRYPNGIGVPLPENDVYNKIANNAAB--TTYGDLTN 478
QY 432 PFDYTKEPSKNIYTDNERKKFNEIKEKIKIEKKKIESDKSYEDRSKSLNDITKYEKL 491
Db 479 P-DNTPITLEDLATNEQARKNLKAIKKKTEAEKNEKLESLKTNYNDKLASFNOOKAPKEA 537
QY 492 LNEIYDSKNNNIDLTNFEKMMGKRYKYVEKLTHTHTFASYSKHNLEKLTALKYME 551
Db 538 AKLFYESKFGNKLTSDFEKFKTQRTQTEYMNKKTLENCL--YGNTKQLSKLNKQLNYLQ 595
QY 552 DYSLRNIVVEKELKYKNIISKIENIETLVENIKKDEBQLEFKKTKDKENRDEKILEV 611
Db 596 DYSLRKDILSDEIEYFSNKKKGLQYNINRLAEAVQAKQNILVA-----SKDVPL 644
QY 612 SDIVKVOQKVLMMNKIDELKKTOLLKKNVELKHNHVPNSYKQENKQBPYYLIVLKKEI 671
Db 645 STLVELIQIKSLTLTKQIEQLNKTLSLRQAQLDKVYVPKSYGNGGKPEPYYLIAVREV 704
QY 672 DKLKVPMPKVESLINEEKNIKTE-----GOSDNSEPESTEGEITGQATTKPGQQAGSAL 725
Db 705 DRLAQFIPAKIENNIKAKE--KTEQVPVVTGESEETS--SVSTEVSVAQSSQSTSTVPA 761
QY 726 EGDS-----VOAQAOEQKQAO-----PVPVPVPEAKAQVPTP-----P 759
Db 762 AGATSPVTVPVTEEAQSSQNAPTTAATPATTPEAATTAATPATTPEAATTTSTTTSTS 821
QY 760 APNNKTEVNSKLDYLEKLYEELNTSYICHKYLIVSHSTMNEKILKOYKITTKEES--KL 817
Db 822 TTTSTTTPVMTKLYLEKLOKFLVFSCHKYLILLQNSTINKDALSKEYALTTEEDKIRTL 891
QY 818 SSCDPLDLLFNQNNIPVMYSMFDSLNLSQLFMFIYEKEMVCNLYKLKD-NDKIKNLL 876
Db 882 KRCELDILLAIONNMPTWLYSIESIVDGLQNIYAELYEKEMHYIYKLDENPSIKSIL 941
QY 877 EBA-----KYSTSVKTSLSSSS 893
Db 942 VRAGVIDPEFVASPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1001
QY 894 MQPLSLTPQDKPEVSANDDTSHSTLNLSLKLNFENILSLGKNKIYQELIGOKSS----- 948
Db 1002 GASASPTASNPGASASSTPTPAAASNPGASASSTPTASNPGASASSTPTQVTPQVPA 1061
QY 949 -----ENFYEKILKSDTFYNESTFNVS 973
Db 1062 PAAPAAVPAQPADSNGSITTRAESIEDIPADDFELDNLYKSLQRIDG-NNTEFINFIKS 1120
QY 974 KADDINLSNDESKRKKLEEDINKLTKTLQSLFDPLYKYLKLERFDKKTKVCKYKMQIK 1033
Db 1121 KKELIKALTPK-KVNOLYLEIAHLKELSHYDRIYKYLKLERLEKHEQIQVSNRQIR 1179
QY 1034 KUTLLKEQLSKLNSLNPNKHLONFESVFNNKKEAIEAETENLTENTKILLKHYKGLVK 1093
Db 1180 ELSILKARLLKRRKONINGIFYILSGYVNFENKRRREADQYVDNALKNNDMLLYKARIK 1239
QY 1094 YINGESSPLKLTSEESIOTEDNYASLENFKVLSKLEBKLDKNLNLKPKKLSYLSGSLHHL 1153
Db 1240 YFTSEAVPLKTLTKASLDRETNLYLKEKFRAYSXRELRFKKNINLKESIYSVGGLYHV 1299
QY 1154 IAEKLVINKNVTGNSPSENNTDNNALLESYKFLPEG-----TDVAT-- 1197
Db 1300 FEEFKLLKNKNTGKTNPDTPVEVNAFEQYKELLPGVGTASASPAATAATPTTSADAATQ 1359
QY 1198 -----VVSESGSDTLEQSQPKKPASTHVGAESNTITTSQNVDDVDDVVIIPIGESE 1250
Db 1360 RATPESGSGSGSGSVSVSTPEEVAR--GSGENAVVSGSSVDD----- 1402
```


Db 1178 SVLSRLKRRKYYINGTFYVLSGFANFNKRREAEQYVDNAIKNTDMLLKYKARSYF 1237
Qy 1096 NGESSPLKTLSEESIOTEDNVASLENKSVLSKLGKLDNLNLEKFKKLISYLSGLHLIA 1155
Db 1238 TSEAVPLKTLKSIDREANYLKEFRAYSRLRLKKNLNKGRITTVSGGLHVF 1297
Qy 1156 ELKEVIKNKNTYGNPSNNNTVNNALESYKFLPEGTD-----VATVSESGSDTLEQS 1210
Db 1298 EFKELKNKGTGTNPENAPEVKAPEQYKELLPGATTAPVVPVAPAPATAAPAA 1357
Qy 1211 QPKP-----ASTHVGAESNTITTSQNVDDVDVVIIPFGESEEDYDGLQ 1258
Db 1358 DAPVAAAAAASGSAATEGAATVTVVASSND-----DDDDMDQ 1402
Qy 1259 VVTGEAVTPSIDNILSKIENEYEVLYLKLPLAGVYKSKOLENNVMTFNVNVKDIINSR 1318
Db 1403 IANAQS--TDEVKDILDFAKSENEYIYKSLGNTYKSKFKHMLKEFSMIKEDINTGLNYK 1461
Qy 1319 FNKRENKVNLESILIPYKDLTSSNVVVKDYKPLNKEKRDKFLSSYNYKDSITDINF 1378
Db 1462 LEKRNDFDLVSYELAFKDLTNTKFNKPNPYQLLDNDKDKQMLNKYAIKGVTEDET 1521
Qy 1379 ANDVLGYKYLSEKYSKSDLSIKYI-----NDKQGENEYKLPNLTLYKTVDNKID 1433
Db 1522 ATDGIIEFNKMIELYKPOLNAVNDQIAAIGTEPTDAEKKYAPIFEDLKLGLYETILNGAE 1581
Qy 1434 LFVTHLEAKVLYNYEYKSNVEVKIKELNYLKTIOQKLADF----KNNNFVGIADLSTDY 1489
Db 1582 EFSELLQHKLENYKIERAGPDILMANLEYIRIDEKLEDFVESAEKKNHTASIA-----1635
Qy 1490 NHNLLKFLSTGMVFENLAKTVLSNLDGNLQGLMNI--SOHQCVKQKOPNSGCFRHL 1548
Db 1636 -----LNNLNSGLVTEGESKKILAKMLNDAMDLLGTSGNHVHCISTSTPDNAGCFRYDD 1690
Qy 1549 ERECKCLLNYKQ--EGDKCVENPNPTCENNNGGCDADAKTEEDSGSN--GKKITCECT 1604
Db 1691 GTEWRLLGLGKDDGNRCVADAPVCNNNGGCDNADCREVENTDRDPSKKIVCTCK 1750
Qy 1605 KPDSPYLFDDGIFCSSNFGISFLILMLILYS 1637
Db 1751 EPNNAVYAGVFCSSGFMGLSILLITLIVEN 1783

RESULT 20

Q25668 ID Q25668 PRELIMINARY; PRT; 1766 AA.
AC Q25668;
DT 01-NOV-1996 (T-REMBLrel. 01, Created)
DT 01-NOV-1996 (T-REMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-REMBLrel. 19, Last annotation update)
DE Merozoite surface protein-1 precursor.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AS;
RX MEDLINE=94187794; PubMed=75111214;
RA McKean P.G., O'Dea K., Brown K.N.;
RT "Nucleotide sequence analysis and epitope mapping of the merozoite
surface protein 1 from Plasmodium chabaudi chabaudi AS.";
RL Mol. Biochem. Parasitol. 62:199-210(1993).
DR EMBL; L22982; AAA29499.1; -;
KW Signal.
FT SIGNAL.
FT CHAIN 1 19 POTENTIAL.
SQ SEQUENCE 1766 AA; 197557 MW; 6CFE98E26EDBD4C CRC64;

Query Match 29.6%; Score 2494.5; DB 5; Length 1766;
Best Local Similarity 32.0%; Pred. No. 2.1e-70;
Matches 603; Conservative 334; Mismatches 581; Indels 365; Gaps 43;

Qy 1 MKIIFLCSLFFIINTQCVTHESYQELVKKLEALEDAVLGTYSLSLFOKRMVNLNEGTS 60

Db 1 MKAIGLFSFVFFAIYCKSETIGVYNDLVHKLLELSVEGLFQKSOVIVNAQSPT 60
Qy 61 AVTTSTPGSKSVASGGSGSVASGGSVASGGSVASGGSGNSRRTPSPNSSDS 120
Db 61 -----PVDPTDS 68
Qy 121 DAKSYADLKHVRNRYLLTIKELYPQLFDLTNHHMLTLCODNIHGFKYLDGYEENELLYK 180
Db 69 -----NFAPKLOEITDFEELGTEQKELVNLKLGPNKYGLKYLESKEEFNELMA 122
Qy 181 LNFYFDLLRAKLDVANDYCOIPFNKIRANELDVLKLVFGYRKPLDNLKDNVGMED 240
Db 123 INFYDVLDRDLNDMCANNYCEIPEHLKINVEIEMLKVVGLYRKPTENIQDVLKLEE 182
Qy 241 YIKKNTKIENINELI-----EESKVTIDKNKNATKEEKKKLYQAOVDSIYNKQ 291
Db 183 YIANKKATATLNNLIETEAKITPEEAECDNTCTNDKYKKKAIQAMVNYFYKKQ 242
Qy 292 LEEAHLISVLEKRDITLKKNNIKELDKINEKNP-----328
Db 243 LAEIKKVIEVLEKRVATLKKNEAIKPLLOQIEAIRGPATVTEGQITTEGTEETKQNDAA 302
Qy 329 -----PPANSNT-----PNTLL-----DNKKI 347
Db 303 QTATKTTTTGSGAGANTNETANAGOTATVTAGTETNTVTVVPPVPLTEEKAKKI 362
Qy 348 EEHEKEIKEIAKTIKFNIDSLSFTDPLEEYVLR--KKNIDISA---KVETKESPEP--N 401
Db 363 AELAQIKEIAKTIKFNLDGIFVDVLELEYKKEKKNESCHSTSSCHNKKPTETVPLNV 422
Qy 402 EYPNGVTYPLS-----YNDINNALNELNSFGDLINPFDYTKEPSKNIYTDNE--RKKFINE 455
Db 423 RYPNGISYPLTEEVYVYSKIAHNAAE--TTYDGLTN--VDNT--PITEDLTNEQARKNLKA 478
Qy 456 IKEKIKTEKKIESDKSYEDRSKSLNDITKEYEKLNEIYDSKFNNDIDLTNPEKMGK 515
Db 479 IKKIEAEQKLVTKADYDTKLAEFNGQKTPFKEAAKNFESFRNKLTTETFEAFKTK 538
Qy 516 RYSYKVEKLTHNTFASYENSKHLEKLTALKYMEDYSLRNIVVEKELKYKNLISKIE 575
Db 539 RTEYMKKAA--LVCEYGNTOQLINKLQNLVLDYGLRKEIVNTEIEVFSNKKSELO 596
Qy 576 NEIETLVENIKKDEQLPEKKITKDNKPDDEKILEVSDIVKQVQVLLMMKIDELKTKQ 635
Db 597 YNINRLANVAQAKNILVA-----SKHIPSLTLVELQIQKSLTLKLEQNKTE 645
Qy 636 LILANVELKHNHVPNSYKQENKOEPPYVYLIVLKKKIDKLKVPMPKVESLINEEKNIKTE 695
Db 646 FSLNKAHLKDKIYVPQTYGKKGPEPPYVYLIAIKKEIDRLAKFIPKIDDMIEKOKMEOE 705
Qy 696 ----GQSD-----NSEPSTGEITGQATTKPGQAGSALSGDSVQAQAOEQ-----737
Db 706 PVATGESEQVTTSSGTVTSTQSAQAPTTPAAPAPAKAEGAETTEKAKEDITETKEPTAT 765
Qy 738 -----KQAP--PVPVPVPEAKAQVPTPPAPVNNKNTENYSKLDYLEKL 778
Db 766 EETSETPAAAPETTPVTEEAQAPAAEVPAPTTPAAPAPAAAPAK---PVMTKLYLEKL 822
Qy 779 YEPINTSYICHKYLVSHSTWNEKILQYKITEEES--KLSSCDPLDLLFNQNNIPVM 836
Db 823 KKFLAFSYACHKYVLLQNSTINKDALSKYALTPEEDKIRTLKRCSELDVLLAIONMPTM 882
Qy 837 YSNFDSLNSLSQLFMEIYEKEMCNLYKLDK--NDKIKNLL-----876
Db 883 YSLYENVVDGLQNIYTELEKEMMYHIYNLKKDPAVKALLVKAAGVIEPEPEPVPPTPAV 942
Qy 877 -----EAKKVTSTVKTLSSSMOPLSLTPQDKPEVSAND-----912
Db 943 PAPETAPVQEAQAQETTOETTPGTTTTPETATPGTTTTPETTPAEPNASETVTQEGTTP 1002
Qy 913 -----TSHSTNNLSLKFENIL 930

```
Db 1003 AEAPAAQKQAGASAPSAETVPAPAPAPQPVTSQASTQVSGQSTNGEGNTKV----- 1057
QY 931 SLGKNKIYQELIGQKSSSENFYEKILKSDTFYNESFTNFVKSABDINSLNDESKRKL 990
Db 1058 ---SAESEDEMFDDPEVDNFYSYLQVDG--NTQFIDIFIKSKKLINALTP--KVNL 1112
QY 991 BEDINKLKTQLSFDLYNKYLKRLRFDKKTGVGYKMQIKKLTLLKEQLESKLSLN 1050
Db 1113 YLIDIAHLKLESHYNYRYKYKLRLYOKHEQIEAANKVKEISVLKSRLLKRYKYN 1172
QY 1051 NPKHVLQNFVFNKKEAEIAETENTLTKLLKHXYGLVYKNGESSPLKTLSEESI 1110
Db 1173 GTFYVLSGFANFNKREAEQYQVNAIKNTDMLLKYYKARSKYFTSEAVPLKTLTKTSI 1232
QY 1111 QTEDNYASLKNFLSKGLKDLNLEKKKLSYLSLSSGLHLIAELKEVIKNKNTGNS 1170
Db 1233 DREANYLKIEKFRAYSRLELRLLKKNINLKGERTVYVSGGLHVFEEFKELLANKYTKT 1292
QY 1171 PSENNTVNNALSKYKFLPEGTD-----VATVSESGSDTLEQSQP-----KKPAS- 1217
Db 1293 NPENAPEVIKAPQYKELLKPGATTAPVPTPAVAPASATTPADEPAAAAAASG 1352
QY 1218 --THVGAESNT--ITTSQNVDDVDVLIIVPIFGESEEDYDLGQVVTGEAVTPSVI 1273
Db 1353 SATTTGGGCTPVANSNDOD-----DDDDMDQIANAQS--TDGEVKDI 1396
QY 1274 LSKINEVEVLYKPLAGVYKSLKQLENNVMTFVNNVKDI---LNSRFNKNREKNVLE 1330
Db 1397 LDEFKSENEYIVTKSLGNTYKSEKHM---LKEFSMIKEDIITGLNYKLEKRNDFDLVLS 1453
QY 1331 SDLIPYKDLTSSNYVVKDPYKFLNKEKDKFLSSVNYTKDSIDTIDINPANDVLGYKILS 1390
Db 1454 YELALFKDINTKVFKNPYQLLNDKDKQMINLKAYKAGVTEDTETATDGIIEFNKMV 1513
QY 1391 EYKYSDDL-----SIKKYINDKOGENEKYLPLFNNLETLYKTVNDKIDLFVHLEAKV 1443
Db 1514 ELYKPOLNAVNEQIAIEKETDKE--EKKYVPFEDLKGUYETILNGAEFSELLQHL 1572
QY 1444 LNYTEKSNVEVKTELNYLKTIOKLADE----KKNNFVGIADLSTDYNNHNLTKFL 1499
Db 1573 ENYKIEKAGFDILMANLETYIRIDEKLEDFVESAEKNKHASIA-----LNNLN 1621
QY 1500 STGMVFENLAKTVLSNLDGNLQGLMNT--SOHQCVKQKQCPNSGCFPHLDERBECKCLLN 1558
Db 1622 KSLGVTGESKKILAKILNDAMDLLGIGSNHVCIGTNIPEAGCFYDGGKEWRCLLG 1681
QY 1559 YKQ--EGDKCVENPNPTCNENNGGCDADAKCTEEDSG--SNGKKITCECTKPDSPYPLFDG 1614
Db 1682 FKNNDGTRCEKDTEATCSINKGCDPSAECREVENADRENSKRVCTCPTNAYYDG 1741
QY 1615 IFCSNLFGLISFLILLMLILYS 1637
Db 1742 VFCSSGFMGLSILLITLIVFN 1764
```

RESULT 21

Q25924

ID Q25924 PRELIMINARY; PRT; 651 AA.

AC Q25924;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Merozoite surface antigen 1 (Fragment).

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RO-71;

RX MEDLINE=92275047; PubMed=1592091;

RA Olafsson P., Matile H., Certa U.;

RT "Plasmodium falciparum: the repetitive MSA-1 surface protein of the

RT RO-71 isolate is recognized by mouse antibody against the

RT

RL nonrepetitive repeat block of RO-33.;"

RN Exp. Parasitol. 74:381-389(1992).

RP [2]

RC SEQUENCE FROM N.A.

RX STRAIN=RO-71;

RX MEDLINE=95354793; PubMed=7628566;

RA Tolle R., Bujard H., Cooper J.A.;

RT "Plasmodium falciparum: variations within the C-terminal region of

merozoite surface antigen-1.;"

RL Exp. Parasitol. 81:47-54(1995).

DR EMBL; Z35329; CAA8458.1.;

DR InterPro; IPR000561; EGF-like.

DR Pfam; PF00008; EGF; 1.

FT NON_TER 1

SQ SEQUENCE 651 AA; 74134 MW; AA2137E699255150 CRC64;

Query Match

Best Local Similarity 25.5%; Score 2146; DB 5; Length 651;

Matches 422; Conservative 64.5%; Pred. No. 5.6e-60;

Mismatches 82; Indels 114; Gaps 36; Gaps 5;

QY 1019 FDKKTKVGYKMQIKKLTLLKEQLESKLSLNPNKHVLFQNFVFNKKEAEIAETENTL 1078

Db 1 FNKKKELGQDKMQIKKLTLLKEQLESKLSLNPNVFNQNFVFNKKEAEIAETENTL 60

QY 1079 ENTKILLKHXYGLVYKNGESSPLKTLSEESIQTEDNYASLENFVLSKLEGLKDLNML 1138

Db 61 ENTKILLKHXYGLVYKNGESSPLKTLSEESIQTEDNYANLEKFRALSKIDKLDNHL 120

QY 1139 EKKLSYLSLSSGLHLIAELKEVIKNKNTGNSPSNNTDNNALSKYKFLPEGTDVATV 1198

Db 121 GKXKLSLSSGLHLITELKEVIKNKNTGNSPSNNTDNNALSKYKFLPEGTDVATV 179

QY 1199 VSE-----SGSDTLEQSQKPKPASTHVGAESNTITTSQNVDDVDVLIIVP 1244

Db 180 VTPQPDVTPSPSVRVSGSGSTKEETQIPTSGSLTLEQVQVQLQNYDEEDSLVLP 239

QY 1245 IFGESEEDYDLGQVVTGEAVTPSVIDNLSKIENEVEVLYKPLAGVYKLSKQLENNV 1304

Db 240 IFGESEEDYDLGQVVTGEAVTSV--MDNILSGFENEVDVYLYKPLAGVYKLSKQLENNV 298

QY 1305 MTFVNVKDLINSRNKRNFKNVLESILPYKDLTSSNYVVKDPYKFLNKEKDKPLSS 1364

Db 299 FTFNLNLDILNSRLKRYKFLDVLSDLMQFKHISSEYIIESFKLLNSEQNTLKS 358

QY 1365 YNVIKDSIDTIDINPANDVLGYKILSEKYKSDLSIKKYI-----N 1405

Db 359 YKIKESVENDIKFAQEGISYIEKVLAKYKDLLESIKKVIKEKEEFPSPPTTSPAKT 418

QY 1406 DKOGENEKYLPFLNNIETLYKTVNDKIDLFVHLEAKVLYTEKSNVEYKIKELNYLKT 1465

Db 419 DEQKESKFLPFLNTIETLYNNLVNKIDDLYLINLAKINDCNVEKDEAHVKITKLSDLKA 478

QY 1466 IQDKLADFKNKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLDGNLQGLM 1525

Db 479 IDDKIDLKFNHNDFEAKLLNDDTKKMLGKLLSTGLV--QNPNTIISKLEGGKFDML 537

QY 1526 NISQHCYKQCPQNSGCFPHLDERBECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 1585

Db 538 NISQHCYKQCPQNSGCFPHLDERBECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 597

QY 1586 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSSNPLGLISFLILLMLILYSFI 1639

Db 598 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSSNPLGLISFLILLMLILYSFI 651

RESULT 22

Q25923

ID Q25923 PRELIMINARY; PRT; 652 AA.

AC Q25923;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Merozoite surface antigen 1 (Fragment).

OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 1-298 FROM N.A.
RC STRAIN=FCH5/NF7;
RX MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens of
Plasmodium falciparum";
RL Mol. Biochem. Parasitol. 27:291-302(1988).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=FCH5/NF7;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
merozoite surface antigen-1";
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL; 235328; CAA84557.1; -;
DR InterPro: IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 1
SQ SEQUENCE 552 AA; 74292 MW; 2B6A87737B490A62 CRC64;

Query Match 25.2%; Score 2119.5; DB 5; Length 652;
Best Local Similarity 63.8%; Pred. No. 3.8e-59;
Matches 418; Conservative 84; Mismatches 116; Indels 37; Gaps 5;

QY 1019 FDKKTVGKVMQIKKLTLLKEQLSKLSNPNPKHVLPNSVFFNKKKAEIAETNTL 1078
Db 1 FNKKELGQDKMQIKKLTLLKEQLSKLSNPNPNHVNLPNSVFFNKKKAEIAETNTL 60

QY 1079 ENTKILLKHVKGLVYKNGESSPLKLTSESIOTEDNYASLENFKVLSLEGKLDKNL 1138
Db 61 ENTKILLKHVKGLVYKNGESSPLKLTSESIOTEDNYANLEKFRVLSKIDGKLDNHL 120

QY 1139 EKKKLSYSLGHLHIAELKEVKNKNTGNSPSENNTDYNNALESYKFLPEGTVDATY 1198
Db 121 GKKLSPSLGHLHIAELKEVKNKNTGNSPENKKVNEALKSVENFLPE-AKVTTV 179

QY 1199 VSE-----SGSDTLEQSQPKPKPASTHVGAESENTTTSQNVDEDDVVIIP 1244
Db 180 VTPPPQDVPSPLSVRVSGSGSTKEETQIPTSGSLLELQVQVQSQNYDEEDDSLVLP 239

QY 1245 IFGESEDDYDLGQVVTGEAVTPSVIDNLSKIENEYEVLYLKLPGVYRSLLKOLENV 1304
Db 240 IFGESEDDYDLGQVVTGEAVT-MDNILSGFENEYEVLYLKLPGVYRSLLKQIEKNI 298

QY 1305 MTFNVNPKDILNSRFNKNFKNVLESDLIPYKDLTSSNVVVDKPYKFLNKRDRFLSS 1364
Db 299 FTENLNLDILNSRLKRRKFLVDYLEDLMQFKHISSEYIIEDSPKLLANSEQKNTLLKS 358

QY 1365 YNTIKSIDTDINFANDVGLYKILSEKYSKSDLSIKKYI----- 1404
Db 359 YKIKESVENDIRFAQEGISYEKVLAKYKDDLESIKKVIKEKEKPPSPPTTPSPAK 418

QY 1405 NDKOGENEKYLPLNNIETLYKTVNDKIDLFVIHLEAKVLNYYEKSNNVVKIKELNYLK 1464
Db 419 TDSQKESKELPELTNIETLYNLVNNKIDYILNLAKINDCNVCEKDEAHVKTIKLSDLK 478

QY 1465 TIQDKLADFKNKNFVGIADLTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQGM 1524
Db 479 AIDKIDLFKNPYDFAIKKLINDTKKMLGKLLSTGLV-QNFPNTIISKLEEGFQDM 537

QY 1525 LNTSQHOCVKKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDAD 1584.
|||||

Db 538 LNTSQHOCVKKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDAD 597
QY 1585 AKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSNPLGIFSLLILMLILYSFI 1639
Db 598 ATCTEKDGSGRKAITCECTKPDSPYPLFDGIFCSSNPLGIFSLLILMLILYSFI 652

RESULT 23
Q03999 PRELIMINARY; PRT; 400 AA.
ID Q03999;
AC Q03999;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 precursor (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RN SEQUENCE FROM N.A.
RA Blackman M.J., Ling I.T., Nicholls S.C., Holder A.A.;
RT "Proteolytic processing of the Plasmodium falciparum merozoite surface
protein-1 produces a membrane-bound fragment containing two epidermal
growth factor-like domains";
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; M64681; AAA29709.1; -;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 24 POTENTIAL
FT CHAIN 25 286 MEROZOITE SURFACE PROTEIN 1.
FT CHAIN 287 400 MEROZOITE SURFACE PROTEIN 1.
SQ SEQUENCE 400 AA; 45824 MW; 537F075058626AC2 CRC64;

Query Match 24.9%; Score 2101; DB 5; Length 400;
Best Local Similarity 100.0%; Pred. No. 8.8e-59;
Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1240 VIIVPIFGESEDDYDLGQVVTGEAVTPSVIDNLSKIENEYEVLYLKLPGVYRSLLKQ 1299
Db 1 VIIVPIFGESEDDYDLGQVVTGEAVTPSVIDNLSKIENEYEVLYLKLPGVYRSLLKQ 60.

QY 1300 LENNVMTFNVNVDILNSRFNKNFKNVLESDLIPYKDLTSSNVVVDKPYKFLNKEKR 1359
Db 61 LENNVMTFNVNVDILNSRFNKNFKNVLESDLIPYKDLTSSNVVVDKPYKFLNKEKR 120

QY 1360 KFLSSYNYIKDSITDINFANDVGLYKILSEKYSKSDLSIKKYINDKOGENEKYLPLN 1419
Db 121 KFLSSYNYIKDSITDINFANDVGLYKILSEKYSKSDLSIKKYINDKOGENEKYLPLN 180

QY 1420 NIETLYKTVNDKIDLFVIHLEAKVLNYYEKSNNVVKIKELNYLKTIODKLADFKNKNF 1479
Db 181 NIETLYKTVNDKIDLFVIHLEAKVLNYYEKSNNVVKIKELNYLKTIODKLADFKNKNF 240

QY 1480 VGTADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQMLNLSHQCVKKQCPQ 1539
Db 241 VGTADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQMLNLSHQCVKKQCPQ 300

QY 1540 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADACTEDSGSNKKI 1599
Db 301 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADACTEDSGSNKKI 360

QY 1600 TCCTKPDSPYPLFDGIFCSSNPLGIFSLLILMLILYSFI 1639
Db 361 TCCTKPDSPYPLFDGIFCSSNPLGIFSLLILMLILYSFI 400

RESULT 24
Q9BMG8

```

ID Q9BMG8 PRELIMINARY; PRT; 376 AA.
AC Q9BMG8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCB-1;
RA Li X.R., Goel V.K., Liu S.C., Chishti A.H., Oh S.S.;
RT "42 kDa subfragment of MSP-1 gene of Plasmodium falciparum FCB-1 isolate."
RL Submitted (DRC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF325919; AAK07641.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 376 AA; 43218 MW; EB2A72EDF231A5AF CRC64;

Query Match 23.5%; Score 1979; DB 5; Length 376;
Best Local Similarity 100.0%; Pred. No. 5.2e-55;
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1264 AVTPSVINILSKITENEVEVLYLPLAGVYRSLKQLNNVMTFNVVVKDILNRFNFKRE 1323
Db 1 AVTPSVINILSKITENEVEVLYLPLAGVYRSLKQLNNVMTFNVVVKDILNRFNFKRE 60

Qy 1324 NFKNVLESDLIPYKDLTSSNVVVDYPYKFLNKRKDRKFLSSYNYIKDSIDTDFNFANDVL 1383
Db 61 NFKNVLESDLIPYKDLTSSNVVVDYPYKFLNKRKDRKFLSSYNYIKDSIDTDFNFANDVL 120

Qy 1384 GYYKILSEKYSDDLSDISKYKINDQGENEKYLPFLNNIETLYKTVDNKIDILFVHLEAKV 1443
Db 121 GYYKILSEKYSDDLSDISKYKINDQGENEKYLPFLNNIETLYKTVDNKIDILFVHLEAKV 180

Qy 1444 LNYTEKSNVEVKIKELNYLKTIDQKADFPKKNFVGIADLTSDYDNNHLLTFLSLGGM 1503
Db 181 LNYTEKSNVEVKIKELNYLKTIDQKADFPKKNFVGIADLTSDYDNNHLLTFLSLGGM 240

Qy 1504 VFENLAKTVLSNLLDGNLQGLMNI SOHCYVKKQCPQNSGCFRHLDERECKLLNYKQEG 1563
Db 241 VFENLAKTVLSNLLDGNLQGLMNI SOHCYVKKQCPQNSGCFRHLDERECKLLNYKQEG 300

Qy 1564 DKCVENPNTCNENGGCDADAKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSNFL 1623
Db 301 DKCVENPNTCNENGGCDADAKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSNFL 360

Qy 1624 GISFLILMLILYSFI 1639
Db 361 GISFLILMLILYSFI 376

RESULT 25
Q9NAT3
ID Q9NAT3 PRELIMINARY; PRT; 360 AA.
AC Q9NAT3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC GOEL V.K., Liu S., Chishti A.H., Oh S.S.;
RT "38 kDa subfragment of MSP-1 gene of FCB-1 strain of Plasmodium falciparum."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF286876; AAF67595.1;
FT NON_TER 1
SQ SEQUENCE 360 AA; 40993 MW; 9A5E9DE65E5A680B CRC64;

Query Match 21.6%; Score 1818; DB 5; Length 360;
Best Local Similarity 99.7%; Pred. No. 5.2e-50;
Matches 359; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 902 QDKPEVSANDDTSHSTNLNLSLKLFENILSLGKKNKIYOELIGOKSSSENFYKILKSDST 961
Db 1 QDKPEVSANDDTSHSTNLNLSLKLFENILSLGKKNKIYOELIGOKSSSENFYKILKSDST 60

Qy 962 FYNESFTNFVSKADDDINSLNDESKRKKLEEDINKLKTQLSFDLYNKYKLERLDFDK 1021
Db 61 FYNESFTNFVSKADDDINSLNDESKRKKLEEDINKLKTQLSFDLYNKYKLERLDFDK 120

Qy 1022 KKTGKYKMQIKKLTLLKEQLESKLSLNNPKHVLQNFVFNKKKEAEIAETENTLENT 1081
Db 121 KKTGKYKMQIKKLTLLKEQLESKLSLNNPKHVLQNFVFNKKKEAEIAETENTLENT 180

Qy 1082 KILKHYKGLVYKNGESSPLKTLSESIQTDNTASLENFKVLKLGKLNKLNLEKK 1141
Db 181 KILKHYKGLVYKNGESSPLKTLSESIQTDNTASLENFKVLKLGKLNKLNLEKK 240

Qy 1142 KLSYSSGLHHLIAELAEVIRKNTGTGNSPENNTDNNALSKYKFLPEGTDVATVYSE 1201
Db 241 KLSYSSGLHHLIAELAEVIRKNTGTGNSPENNTDNNALSKYKFLPEGTDVATVYSE 300

Qy 1202 SGSDTLEQSQPKPASTHVGAESNTITTSQNVDDVDVVIIPFGESEEDYDDLGVVVT 1261
Db 301 SGSDTLEQSQPKPASTHVGAESNTITTSQNVDDVDVVIIPFGESEEDYDDLGVVVT 360

RESULT 26
Q25975
ID Q25975 PRELIMINARY; PRT; 569 AA.
AC Q25975;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 21.1%; Score 1778; DB 5; Length 569;
Best Local Similarity 60.8%; Pred. No. 1.5e-48;
Matches 348; Conservative 79; Mismatches 109; Indels 36; Gaps 5;

Qy 1101 PLKTLSESIQTDNTASLENFKVLKLGKLNKLNLEKKLSYSSGLHHLIAELKEV 1160
Db 1 PLKTLSESIQTDNTASLENFKVLKLGKLNKLNLEKKLSYSSGLHHLIAELKEV 60

Qy 1161 IKKNYTGNSPENNTDNNALSKYKFLPEGTDVATVYSE-----SGSDT 1206
```

```

ID Q9BMG8 PRELIMINARY; PRT; 376 AA.
AC Q9BMG8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCB-1;
RA Li X.R., Goel V.K., Liu S.C., Chishti A.H., Oh S.S.;
RT "42 kDa subfragment of MSP-1 gene of Plasmodium falciparum FCB-1 isolate."
RL Submitted (DRC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF325919; AAK07641.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 376 AA; 43218 MW; EB2A72EDF231A5AF CRC64;

Query Match 23.5%; Score 1979; DB 5; Length 376;
Best Local Similarity 100.0%; Pred. No. 5.2e-55;
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1264 AVTPSVINILSKITENEVEVLYLPLAGVYRSLKQLNNVMTFNVVVKDILNRFNFKRE 1323
Db 1 AVTPSVINILSKITENEVEVLYLPLAGVYRSLKQLNNVMTFNVVVKDILNRFNFKRE 60

Qy 1324 NFKNVLESDLIPYKDLTSSNVVVDYPYKFLNKRKDRKFLSSYNYIKDSIDTDFNFANDVL 1383
Db 61 NFKNVLESDLIPYKDLTSSNVVVDYPYKFLNKRKDRKFLSSYNYIKDSIDTDFNFANDVL 120

Qy 1384 GYYKILSEKYSDDLSDISKYKINDQGENEKYLPFLNNIETLYKTVDNKIDILFVHLEAKV 1443
Db 121 GYYKILSEKYSDDLSDISKYKINDQGENEKYLPFLNNIETLYKTVDNKIDILFVHLEAKV 180

Qy 1444 LNYTEKSNVEVKIKELNYLKTIDQKADFPKKNFVGIADLTSDYDNNHLLTFLSLGGM 1503
Db 181 LNYTEKSNVEVKIKELNYLKTIDQKADFPKKNFVGIADLTSDYDNNHLLTFLSLGGM 240

Qy 1504 VFENLAKTVLSNLLDGNLQGLMNI SOHCYVKKQCPQNSGCFRHLDERECKLLNYKQEG 1563
Db 241 VFENLAKTVLSNLLDGNLQGLMNI SOHCYVKKQCPQNSGCFRHLDERECKLLNYKQEG 300

Qy 1564 DKCVENPNTCNENGGCDADAKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSNFL 1623
Db 301 DKCVENPNTCNENGGCDADAKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSNFL 360

Qy 1624 GISFLILMLILYSFI 1639
Db 361 GISFLILMLILYSFI 376

RESULT 25
Q9NAT3
ID Q9NAT3 PRELIMINARY; PRT; 360 AA.
AC Q9NAT3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC GOEL V.K., Liu S., Chishti A.H., Oh S.S.;
RT "38 kDa subfragment of MSP-1 gene of FCB-1 strain of Plasmodium falciparum."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF286876; AAF67595.1;
FT NON_TER 1
SQ SEQUENCE 360 AA; 40993 MW; 9A5E9DE65E5A680B CRC64;

Query Match 21.6%; Score 1818; DB 5; Length 360;
Best Local Similarity 99.7%; Pred. No. 5.2e-50;
Matches 359; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 902 QDKPEVSANDDTSHSTNLNLSLKLFENILSLGKKNKIYOELIGOKSSSENFYKILKSDST 961
Db 1 QDKPEVSANDDTSHSTNLNLSLKLFENILSLGKKNKIYOELIGOKSSSENFYKILKSDST 60

Qy 962 FYNESFTNFVSKADDDINSLNDESKRKKLEEDINKLKTQLSFDLYNKYKLERLDFDK 1021
Db 61 FYNESFTNFVSKADDDINSLNDESKRKKLEEDINKLKTQLSFDLYNKYKLERLDFDK 120

Qy 1022 KKTGKYKMQIKKLTLLKEQLESKLSLNNPKHVLQNFVFNKKKEAEIAETENTLENT 1081
Db 121 KKTGKYKMQIKKLTLLKEQLESKLSLNNPKHVLQNFVFNKKKEAEIAETENTLENT 180

Qy 1082 KILKHYKGLVYKNGESSPLKTLSESIQTDNTASLENFKVLKLGKLNKLNLEKK 1141
Db 181 KILKHYKGLVYKNGESSPLKTLSESIQTDNTASLENFKVLKLGKLNKLNLEKK 240

Qy 1142 KLSYSSGLHHLIAELAEVIRKNTGTGNSPENNTDNNALSKYKFLPEGTDVATVYSE 1201
Db 241 KLSYSSGLHHLIAELAEVIRKNTGTGNSPENNTDNNALSKYKFLPEGTDVATVYSE 300

Qy 1202 SGSDTLEQSQPKPASTHVGAESNTITTSQNVDDVDVVIIPFGESEEDYDDLGVVVT 1261
Db 301 SGSDTLEQSQPKPASTHVGAESNTITTSQNVDDVDVVIIPFGESEEDYDDLGVVVT 360

RESULT 26
Q25975
ID Q25975 PRELIMINARY; PRT; 569 AA.
AC Q25975;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 21.1%; Score 1778; DB 5; Length 569;
Best Local Similarity 60.8%; Pred. No. 1.5e-48;
Matches 348; Conservative 79; Mismatches 109; Indels 36; Gaps 5;

Qy 1101 PLKTLSESIQTDNTASLENFKVLKLGKLNKLNLEKKLSYSSGLHHLIAELKEV 1160
Db 1 PLKTLSESIQTDNTASLENFKVLKLGKLNKLNLEKKLSYSSGLHHLIAELKEV 60

Qy 1161 IKKNYTGNSPENNTDNNALSKYKFLPEGTDVATVYSE-----SGSDT 1206
```

Db	61	IKKNYTGNSPSENKKVNEALKSYENFLPE-AKVTTVVTPQDPVTSPLSVRVSGSG	119
Qy	1207	LEQSQPKKPASTHVGAESNTITTSQNVDVDDVIVPIFGSEEDYDDLQGVVTVGRAVT	1266
Db	120	STKEETQIPTSGSLLTQLQOVVQLQYQVDEEDSLVPLPIFGSEDNDEYLDQVVTGEAIS	179
Qy	1267	PSVIDNLTLSKLENEYEVILYKPLAGVYRSLLKQLENNYMTFNVNVKDTILNSRFKNKRENF	1326
Db	180	VT-MDNLTSGFENEYDIVYLLPLAGVYRSLLKQLENNYMTFNVNVKDTILNSRFKNKRENF	238
Qy	1327	NVLESDLPYKDLTSSNYVVKVDPYKPLNKEKRDKFLSSNYIKDSIDTDINFANDVLGY	1386
Db	239	DVLESDLMQPKHLSSENYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAQEGISY	298
Qy	1387	KILSEKYKSDLSIKYI-----NDKQENKEYLPFLNNITETLYKT	1427
Db	299	EKLVALYKDDLESIKVKEEKEFFSPPTPPSPAKTDEQKESKFLPFLTNITETLYNN	358
Qy	1428	VNDKIDLFVHLEAKVLNITYEKSINVEYKIKELNYLKTLODKLADPKKNNPVGADLST	1487
Db	359	LVNKIDYDYLNLAKKACINDCNVEKDEAHVKITKLSDLKADDKIDILFKNTWDFEAIKKLIN	418
Qy	1488	DYHNHNNLTFLSTGVWFENLAKTIVLSNLNDGLNLQGLNLSIQHCVKKQCPNSGCFRHL	1547
Db	419	DDTKMDLMLKLLSTGLV-QNFPWTIISKLECKFQDMLNLSIQHCVKKQCPNSGCFRHL	477
Qy	1548	DEREECKLLNYKQEGDKCVENPNTPCNNNGCCDADAKCTEEDSGSGNGKKITCECTKPD	1607
Db	478	DEREECKLLNYKQEGDKCVENPNTPCNNNGCCDADAKCTEEDSGSGNGKKITCECTKPD	537
Qy	1608	SYPLFDGIFGFCSSNFGISGIFLLTILMLILYSFI	1639
Db	538	SYPLFDGIFGFCSSNFGISGIFLLTILMLILYSFI	569
RESULT	28		
Q25977		PRELIMINARY;	PRT; 569 AA.
AC	Q25977;		
DT	01-NOV-1996	(TEMBLrel. 01, Created)	
DT	01-NOV-1996	(TEMBLrel. 01, Last sequence update)	
DT	01-MAR-2002	(TEMBLrel. 20, Last annotation update)	
DE	Major merozoite surface protein (Fragment).		
GN	MSPI.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
OX	NCBI_TaxID=5833;		
ON	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93295445; PubMed=8515786;		
RA	Jongwatitwies S., Tanabe K., Kanbara H.;		
RT	"Sequence conservation in the C-terminal part of the precursor to the		
RT	major merozoite surface proteins (MSPI) of Plasmodium falciparum from		
RT	field isolates.;"		
RL	Mol. Biochem. Parasitol. 59:95-100(1993).		
DR	EMBL; D13350; BAA02611.1.;		
DR	InterPro: IPR000561; EGF-like.		
DR	Pfam; PF00008; EGF; 1.		
KW	EGF-like domain; Merozoite.		
FT	NON_TER		
FT	SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;		

	Query Match	Best Local Similarity	Score 17/22	DB 5	Length 569
	Matches 347	Conservative 79	Mismatches 110	Indels 36	Gaps
Qy	1101	PLKLTSESQTEDNYASLENFVLSKLGKLLKLDNLEKKLSYLSGSLHLLIAELKEV	1160		
Db	1	PLKLTSEVSQTEDNYANLEKFRVLSKIDGKLDNLDLHGCKKLSFLSSGSLHLLITELKEV	60		
Qy	1161	IKKNKNTGNSPSENNTDVNNALSESYYKKFLPEGTADVATVYSE	-----SGSDT	1206	
Db	61	IKKNKNTGNSPSENKKNNKVNALKSYENFLPE-AKVTVTVPQDPVTPSPVSRVYSGSG	119		


```
Db 120 STKEETQIPTSGSLTLELQOVVQLQNYDEEDDSLVVLPPIFGESEDNDEYLDQVVTGEAIS 179
QY 1267 PSVIDNILSKINEVEVLKPLAGVYRSLLKQLENNVMTFNVVKDILNSRFNKRENEK 1326
Db 180 VT-MDNILSGFENEVDVIYKPLAGVYRSLLKQLENNVMTFNVVKDILNSRLKRRKRYL 238
QY 1327 NVLESDLPYKDLTSSNYVVKPYKFLNKKRDKFLSSNYIKDSITDITDINFANDVLGY 1386
Db 239 DVLESDLMQFKHISSEYIIEDSFLLNSEQNTLLKSYKIKESVENDIKFAQEGISY 298
QY 1387 KILSEKYSDDLDSIKKIYI-----NDKQGENEKYLPFLNNIETLYKT 1427
Db 299 EXVLAKYKDLLESIKKIYKEEKEFPSPPTTPSPAKTDEQKESKFLPFLNIETLYNN 358
QY 1428 VNDKIDLFVIHLEAKVLYTYEKSNEVKIKELNYLKTIDQLADFKKNNVFGIADLST 1487
Db 359 LVNKKIDYILINLAKINDCNVEKDEAHVAKITKLSDLKAIDDKIDLFKNTDFAIKKLIN 418
QY 1488 DYNHNNLLTKFLSTGMVFENLAKTVLSNLLDGNLQMLNISQHCVKKQCPQNSGCFRHL 1547
Db 419 DDTKMDLGLKLLSTGLV-QNEPNTIISKLEGGKFDQMLNISQHCVKKQCPQNSGCFRHL 477
QY 1548 DERECKCLLNTYKQEGDKCVENPNTCENNNGGCDADAKCTEEDSGSNGKKITCECTKPD 1607
Db 478 DERECKCLLNTYKQEGDKCVENPNTCENNNGGCDADAKCTEEDSGSNGKKITCECTKPD 537
QY 1608 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1639
Db 538 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569
```

```
RESULT 31
Q25983
ID Q25983 PRELIMINARY; PRT; 569 AA.
AC Q25983;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13355; BAA02616.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64491 MW; 57A6B62F72CE885 CRC64;
```

Query Match 21.0%; Score 1767; DB 5; Length 569;
Best Local Similarity 60.5%; Pred. No. 3.2e-48;
Matches 346; Conservative 78; Mismatches 112; Indels 36; Gaps 5;

```
QY 1101 PKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKLLSYLSSGLHLLIAELKEV 1160
Db 1 PKLTSEESIQTEDNYANLEKFRVLSKIDGKLDNLDNLHGLKLSFLSSGLHLLITELKEV 60
QY 1161 IKKNYTGNSPSENNTDVNNALESYKKFLPECTDVAIVYSE-----SGSDT 1206
Db 61 IKKNYTGNSPENNNKKNVNALESYENFLPE-AKVTIVTPPQDPTSPSLVRSVSGSG 119
QY 1207 LEQSQPKPASTHVGAESENTITTSQNVDDVDVIIVPIFGESEEDYDLQGVVTGEAVT 1266
```

```
Db 120 STKEETQIPTSGSLTLELQOVVQLQNYDEEDDSLVVLPPIFGESEDNDEYLDQVVTGEAIS 179
QY 1267 PSVIDNILSKINEVEVLKPLAGVYRSLLKQLENNVMTFNVVKDILNSRFNKRENEK 1326
Db 180 VT-MDNILSGFENEVDVIYKPLAGVYRSLLKQLENNVMTFNVVKDILNSRLKRRKRYL 238
QY 1327 NVLESDLPYKDLTSSNYVVKPYKFLNKKRDKFLSSNYIKDSITDITDINFANDVLGY 1386
Db 239 DVLESDLMQFKHISSEYIIEDSFLLNSEQNTLLKSYKIKESVENDIKFAQEGISY 298
QY 1387 KILSEKYSDDLDSIKKIYI-----NDKQGENEKYLPFLNNIETLYKT 1427
Db 299 EXVLAKYKDLLESIKKIYKEEKEFPSPPTTPSPAKTDEQKESKFLPFLNIETLYNN 358
QY 1428 VNDKIDLFVIHLEAKVLYTYEKSNEVKIKELNYLKTIDQLADFKKNNVFGIADLST 1487
Db 359 LVNKKIDYILINLAKINDCNVEKDEAHVAKITKLSDLKAIDDKIDLFKNTDFAIKKLIN 418
QY 1488 DYNHNNLLTKFLSTGMVFENLAKTVLSNLLDGNLQMLNISQHCVKKQCPQNSGCFRHL 1547
Db 419 DDTKMDLGLKLLSTGLV-QNEPNTIISKLEGGKFDQMLNISQHCVKKQCPQNSGCFRHL 477
QY 1548 DERECKCLLNTYKQEGDKCVENPNTCENNNGGCDADAKCTEEDSGSNGKKITCECTKPD 1607
Db 478 DERECKCLLNTYKQEGDKCVENPNTCENNNGGCDADAKCTEEDSGSNGKKITCECTKPD 537
QY 1608 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1639
Db 538 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569
```

RESULT 32

Q25967

ID Q25967 PRELIMINARY; PRT; 569 AA.

AC Q25967;

DT 01-NOV-1996 (TremBLrel. 01, Created)

DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)

DE Major merozoite surface protein (Fragment).

GN MSP1.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93295445; PubMed=8515786;

RA Jongwutiwes S., Tanabe K., Kanbara H.;

RT "Sequence conservation in the C-terminal part of the precursor to the

RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from

RT field isolates."

RL Mol. Biochem. Parasitol. 59:95-100(1993).

DR EMBL: D13344; BAA02605.1; -.

DR InterPro: IPR000561; EGF-like.

DR Pfam: PF00008; EGF; 1.

KW EGF-like domain; Merozoite.

FT NON_TER 1

SQ SEQUENCE 569 AA; 64435 MW; BE600D346E51304E CRC64;

Query Match 21.0%; Score 1765; DB 5; Length 569;
Best Local Similarity 60.5%; Pred. No. 3.7e-48;
Matches 346; Conservative 79; Mismatches 111; Indels 36; Gaps 5;

QY 1101 PKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKLLSYLSSGLHLLIAELKEV 1160

Db 1 PKLTSEESIQTEDNYANLEKFRVLSKIDGKLDNLDNLHGLKLSFLSSGLHLLITELKEV 60

QY 1161 IKKNYTGNSPSENNTDVNNALESYKKFLPECTDVAIVYSE-----SGSDT 1206

Db 61 IKKNYTGNSPENNNKKNVNALESYENFLPE-AKVTIVTPPQDPTSPSLVRSVSGSG 119

QY 1207 LEQSQPKPASTHVGAESENTITTSQNVDDVDVIIVPIFGESEEDYDLQGVVTGEAVT 1266

Db 120 STKEETQIPTSGSLTLELQOVVQLQNYDEEDDSLVVLPPIFGESEDNDEYLDQVVTGEAIS 179

Qy 1267 PSVIDNLTLSKIENEYEVLYLKPLAGVYRSLSKQLENNVMTFNVNKDILNSRFNKRENFK 1326

Db 180 VT-MDNTLSGFENEYDVIYLLKPLAGVYRSLSKQLENNVMTFNVNKDILNSRLKRRKYFL 238

Qy 1327 NVLESDLIPYKOLTSNYSVYKDPFLNKKRDKFLSSVNYIKOSIDTDINFANDVLGY 1386

Db 239 DVLESDDMQFKHJSSNEYIIESFKLNSSEQNTLLSKYKIKESVENDIKFAQEGISY 298

Qy 1387 KILSEKYSKDLDSIKKYI-----NDKQGENEYKLPFLNNIETLYK 1426

Db 299 EKVLAKYKODLESIKKVIKEKEKFPSSPTTPSPAKTDEQKESKFLPFLNTIETLYN 358

Qy 1427 TVNDKIDLFVHLEAKVNTYKESVNEVYKIELNLYKTIOQKLADFKNKNNFVGIADLS 1486

Db 359 NLVKNIDYILNLKAKINDCNVEKAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLI 418

Qy 1487 TDYNNHNLTKFSTGMVFENLAKTVLSNLLDGNLOGMLNISOHOCVKKQCPONGCPRH 1546

Db 419 NDDTKMDLKGKLLSTGLV-ONFPNTTISKLEGGFKODMLNISOHOCVKKQCPENGSCPRH 478

Qy 1547 LDRECKCCLLNTKQEGDKCVENPNTCENNGGCDADAACKTEEDSGSGKKITCECTKP 1606

Db 478 LDRECKCCLLNTKQEGDKCVENPNTCENNGGCDADATCTEEDSGSGRKKITCECTKP 537

Qy 1607 DSYPLDFGIFCSSNFGISFLLLMLILYSFI 1639

Db 538 DSYPLDFGIFCSSNFGISFLLLMLILYSFI 570

RESULT 34

Q25982 ID Q25982 PRELIMINARY; PRT: 569 AA.

AC AC Q25982:

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Major merozoite surface protein (Fragment).

GN MSP1.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93295445; PubMed=8515786;

RA Jongutiwes S., Tanabe K., Kanbara H.:

RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP) of Plasmodium falciparum from field isolates."

RL Mol. Biochem. Parasitol. 59:95-100(1993).

DR ENBL; D13354; BAA02615.1; -.

DR InterPro: IPR000561; EGF-like.

DR Pfam; PF00008; EGF; 1.

KW EGF-like domain; Merozoite.

FT NON_TER 1

FT SEQUENCE 569 AA; 64530 MW; 4D734580516D2142 CRC64;

Query Match 20.8%; Score 1753; DB 5; Length 569;

Best Local Similarity 60.1%; Pred. No. 8.e-48;

Matches 344; Conservative 80; Mismatches 112; Indels 36; Gaps

Qy 1101 PLKTLSESIQTEDNYASLENFKVLSKLEGGKLDNLNLEKKKLSYSSGLHHLIAELKEV 1160

Db 1 PLKTLSEVSIQTEDNYANLEKFRVLSKIDGKLDNLHLGKKKLSFLSSGLHHLITELKEV 60

Qy 1161 IKKNYTGNSPSENNTDVNNALSEYKKFLPEGTDTATVYSE-----SGSDT 1206

Db 61 IKKNYTGNSPENNNKVNALKSYENFLPE-AKVTVTVPQDPVTPSPLSRVSGSG 119

Qy 1207 LEQSQPKPASTHVGAESENTITTSQNVDEDDVIIVPIFGSESEDDDLGOVYTGEAVT 1266

Db 120 STEETQIPTSGSLLTEQOVQVOLQNVDEDDSLVLPPIFGSESDNDDEYLOQVVTGEAIS 179

Qy 1267 PSVIDNLTLSKIENEYEVLYLKPLAGVYRSLSKQLENNVMTFNVNKDILNSRFNKRENFK 1326


```
Db 120 STKEETQIPTSGLLTELQOVQVQLQYDEEDSLVVLPIPFGESEDNDEYLDQVVTGEAIS 179
QY 1267 PSVIDNLSKIENEYEVLYLKLPLAGVYRSUKKOLENNVMTFNVVKDILNSRFNKRENF 1326
Db 180 VT-MDNILSGFENEYDVLYLKLPLAGVYRSUKKOLENNVMTFNVVKDILNSRUKKRYFL 238
QY 1327 NVLESOLIPKOLTSNYYVVKDPYKFLNKEKRDQFLSSNYIKDSIDTDINFANDVLGY 1386
Db 239 DVLESOLMQFKHTSSNEYIIEEDSFLLNSQKNTLLSKYIKRESVENDIKFAQEGISY 298
QY 1387 KILSEYKSDLDISIKYI-----NDKQGENEKYLPFLNNIETLYKT 1427
Db 299 EKVLAYKODLESIKKVIKEEFPSPPTTPSPAKTDQKKESKELPFLTNIEIYLYNN 358
QY 1428 VNDKIDLFVHLAKVLTNYEKSNEVYKIKELNYLKTIDQKLADFKNNNFVGIADLST 1487
Db 359 LVNKIDYLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTDFAIKKLIN 418
QY 1488 DYNHNNLLTKFLSTGMVFENLAKTVISNLDGLNQLGMLNISQHCYVKKQCPQNSGCFRHL 1547
Db 419 DDTKKDMLGKLLSTGLV-QNPNTIISKLTIEGKFQDMLNISQHCYVKKQCPNSGCFRHL 477
QY 1548 DERECKCLLNYKQEGDKCVENPNPCNNENGGCDADAKTEEDSGSNGKKITCECTKPD 1607
Db 478 DERECKCLLNYKQEGDKCVENPNPCNNENGGCDADAKTEEDSGSRRKKITCECTKPD 537
QY 1608 SYPLFDGIFCSSNFIIGISFLILMLILYSFI 1639
Db 538 SYPLFDGIFCSSNFIIGISFLILMLILYSFI 569

RESULT 39
Q25865 PRELIMINARY; PRT; 336 AA.
AC Q25865:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91304517; PubMed=1852173;
RA Kerr P.J., Ranford-Cartwright L.C., Walliker D.;
RT "Proof of intragenic recombination in Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 46:185-187(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95107347; PubMed=7808474;
RA Kerr P.J., Ranford-Cartwright L.C., Walliker D.;
RT "Proof of intragenic recombination in Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 66:241-248(1994).
DR EMBL; X52962; CAA37135.1; -.
FT NON_TER 1
FT NON_TER 336
SQ SEQUENCE 336 AA; 37906 MW; 3094DOC5102B73DE CRC64;

Query Match 20.1%; Score 1696; DB 5; Length 336;
Best Local Similarity 96.0%; Pred. No. 3.1e-46;
Matches 334; Conservative 1; Mismatches 1; Indels 12; Gaps 1;

QY 21 THESQELVKKLEALEDAVLTGYSLFQKEKMYLNEGTSGTAVTSTPPSGKSGVASGGSG 80
Db 1 THESQELVKKLEALEDAVLTGYSLFQKEKMYLNEGTSGTAVTSTPPSGKSGVASGG--- 57
QY 81 SVASGGSGVASGGSGVASGGSGNSRRTNPDSNDSDAKSYADLKHVRNYLLTTIK 140
Db 58 -----SGGSGVASGGSGNSRRTNPDSNDSDAKSYADLKHVRNYLLTTIK 108
QY 141 ELKYPOLFDTLNHMLTLCNDIHGFKYLDGYEINELLYKLNFDLLRAKLNDCVANCY 200
```

```
Db 109 ELKYPOLFDTLNHMLTLCNDIHGFKYLDGYEINELLYKLNFDLLRAKLNDCVANCY 168
QY 201 CQIPFNLIKIRANEDLVKLVFGYRKPDLNIDKNVGMEDYIKNKKTIENINELIEESK 260
Db 169 CQIPFNLIKIRANEDLVKLVFGYRKPDLNIDKNVGMEDYIKNKKTIENINELIEESK 228
QY 261 KTDKNKNATKEBEKKLYQAOYDLSYNNKQLEBAHNLSVLEKRIDTFLAKNENIKELLD 320
Db 229 KTDKNKNATKEBEKKLYQAOYDLSYNNKQLEBAHNLSVLEKRIDTFLAKNENIKELLD 288
QY 321 KINEIKNPPANGNPNTLLDKNKKTIEEHEKEIKEIAKTIKFNIDSL 368
Db 289 KINEIKNPPANGNPNTLLDKNKKTIEEHEKEIKEIAKTIKFNIDSL 336

RESULT 40
Q9TZU5 PRELIMINARY; PRT; 356 AA.
AC Q9TZU5:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HN2;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061143; AAC69742.1; -.
FT NON_TER 356
FT NON_TER 356
SQ SEQUENCE 356 AA; 39756 MW; 35042CC13AB890E6 CRC64;

Query Match 19.8%; Score 1670; DB 5; Length 356;
Best Local Similarity 91.6%; Pred. No. 2.1e-45;
Matches 329; Conservative 10; Mismatches 14; Indels 6; Gaps 2;

QY 1 MKIIFFLCSFLFFIINTQCVTHESYQELVKKLEALEDAVLTGYSLFQKEKMYLNEGTSGT 60
Db 1 MKIIFFLCSFLFFIINTQCVTHESYQELVKKLEALEDAVLTGYSLFQKEKMYLNEGTSGT 60
QY 61 AVTTSTPGSK---GSVASGGSGVASGGSGVASGGSGVASGGSGNSRRTNPDSNS 117
Db 61 AVTTSTPGSKSGSGSVASGGSGSVASGGSGVA---SVASGGSGNSRRTNPDSNS 117
QY 118 SOSDAKSYADLKHVRNYLLTIKELYPOLFDTLNHMLTLCNDIHGFKYLDGYEINEL 177
Db 118 SOSDAKSYADLKHVRVONYLFTIKELYPOLFDTLNHMLTLCNDIHGFKYLDGYEINEL 177
QY 178 LYKLNIFYDILLRAKLNDCVANCYQIPFNLIKIRANEDLVKLVFGYRKPDLNIDKNVKG 237
Db 178 LYKLNIFYDILLRAKLNDCVANCYQIPFNLIKIRANEDLVKLVFGYRKPDLNIDKNVKG 237
QY 238 MEDYIKNKKTIENINELIEESKTTDKNNKATKEBEKKLYQAOYDLSYNNKQLEBAHN 297
Db 238 MEDYIKNKKTIENINELIEESKTTDKNNADNEEGKKLYQAOYDLSYNNKQLEBAHN 297
QY 298 LISVLEKRIDTFLAKNENIKELLDKINEIKNPPANGNPNTLLDKNKKTIEEHEKEIKE 356
Db 298 LISVLEKRIDTFLAKNENIKELLDKINEIKNPPANGNPNTLLDKNKKTIEEHEKEIKE 356

RESULT 41
Q9TZU2
```

```

ID O9T202 PRELIMINARY; PRT; 350 AA.
AC O9T202;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HN6.24;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061146; AAC69745.1; -.
FT NON_TER 350
SQ SEQUENCE 350 AA; 39343 MW; ALE3C3AA6AF48FC9 CRC64;

Query Match 19.6%; Score 1649; DB 5; Length 350;
Best Local Similarity 91.0%; Pred. No. 9.4e-45;
Matches 324; Conservative 11; Mismatches 15; Indels 6; Gaps 2;

QY 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEMVNEGTS 60
Db 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEMVNEGTS 60
QY 61 AVTTSTPGSGVTSVSGSGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSG 120
Db 61 AVTTSTPGSGSVT---SGGSVTSVG---SGGSVTSVSG---SGGSVTSVSG 114
QY 121 DAKSYADLKHVRVNYLLTIKELYPQLFDLTNHLMLTCLDNIHGFYKYLIDGYEINELLYK 180
Db 115 DAKSYADLKHVRVNYLLTIKELYPQLFDLTNHLMLTCLDNIHGFYKYLIDGYEINELLYK 174
QY 181 LNFYDILLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNIKDNVGMED 240
Db 175 LNFYDILLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNIKDNVGMED 234
QY 241 YIKKNTTIENINELIEESKTTIDKNKNATKEEKKLYQAOYDLSYNNKOLEEAHNLIS 300
Db 235 YIKKNTTIENINELIEESKTTIDKNKNADNEEGKKLYQAOYDLSYNNKOLEEAHNLIS 294
QY 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTNPNTLLDKNKKIEHEEKEIKE 356
Db 295 VLEKRIDTLKKNENIKKLDKIDEIKNPPANSNGTNPNTLLDKNKKIEHEEKEIKE 350

RESULT 42
Q9T207 PRELIMINARY; PRT; 357 AA.
ID Q9T207;
AC Q9T207;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFA125;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
```

```

RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061141; AAC69740.1; -.
FT NON_TER 357
SQ SEQUENCE 357 AA; 39850 MW; AE9513F521FA9ED5 CRC64;

Query Match 19.4%; Score 1636.5; DB 5; Length 357;
Best Local Similarity 90.6%; Pred. No. 2.4e-44;
Matches 326; Conservative 11; Mismatches 16; Indels 7; Gaps 3;

QY 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEMVNEGTS 60
Db 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEMVNEGTS 60
QY 61 AVTTSTPG---SKGSVASGGSGSVASGSGSVASGSGSVASGSGSVASGSGSVAS 117
Db 61 AVTTSTPGSVASGSGSVASGSGSGSVASG---SGGSVASGSGSVASGSGSVAS 117
QY 118 SDSDAKSYADLKHVRVNYLLTIKELYPQLFDLTNHLMLTCLDNIHGFYKYLIDGYEINEL 177
Db 118 SDSDAKSYADLKHVRVNYLLTIKELYPQLFDLTNHLMLTCLDNIHGFYKYLIDGYEINEL 177
QY 178 LYKLNIFYDILLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNIKDNV 237
Db 178 LYKLNIFYDILLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNIKDNV 237
QY 238 MEDYIKKNTTIENINELIEESKTTIDKNKNATKEEKKLYQAOYDLSYNNKOLEEAH 297
Db 238 MEDYIKKNTTIENINELIEESKTTIDKNKNADNEEGKKLYQAOYDLSYNNKOLEEAH 297
QY 298 LISVLEKRIDTLKKNENIKELLDKINEIK-NPPANSNGTNPNTLLDKNKKIEHEEKEIKE 356
Db 298 LISVLEKRIDTLKKNENIKKLEDDIKTDABEKLTTGSKPNTLLDKNKKIEHEEKEIKE 357

RESULT 43
Q9T208 PRELIMINARY; PRT; 357 AA.
ID Q9T208;
AC Q9T208;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFA12;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061140; AAC69739.1; -.
FT NON_TER 357
SQ SEQUENCE 357 AA; 39850 MW; 498FDF8E2DF1A1D CRC64;

Query Match 19.4%; Score 1634.5; DB 5; Length 357;
Best Local Similarity 90.3%; Pred. No. 2.7e-44;
Matches 325; Conservative 13; Mismatches 15; Indels 7; Gaps 3;

QY 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEMVNEGTS 60
Db 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEMVNEGTS 60
QY 61 AVTTSTPG---SKGSVASGGSGSVASGSGSVASGSGSVASGSGSVASGSGSVAS 117
Db 61 AVTTSTPGSVASGSGSGSVASGSGSVASG---SGGSVASGSGSVASGSGSVAS 117
```

```
Qy 118 SDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCDNIHGFYKYLIDGYEEINEL 177
Db 118 SDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCDNIHGFYKYLIDGYEEINEL 177
Qy 178 LYKLNFFDRLAKLNDVCANDYCOIPENLKITRANELDVLKLVFGYRKPLDNIKDNVYK 237
Db 178 LYKLNFFDRLAKLNDVCANDYCOIPENLKITRANELDVLKLVFGYRKPLDNIKDNVYK 237
Qy 238 MEDYIKKNNKTITANINELIEESKKTIDKNKATKEEKKLYQAOYDLSIYNKQLEEAHN 297
Db 238 MEDYIKKNNKTITANINELIEESKKTIDKNKADNEEGKKLYQAOYDLSIYNKQLEEAHN 297
Qy 298 LISVLEKRIDTLKKNENIKELLDKINEIK-NPPPPANSNGTPTNLDDKNKIEEHEKEIKE 356
Db 298 LISVLEKRIDTLKKNENIKELLDKIDKIDAEKLTGSKPNTLLDKNKIEEHEKEIKE 357
RESULT 44
Q9TZV2 PRELIMINARY; PRT; 360 AA.
AC Q9TZV2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFA9.2;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061136; AAC69735.1; -.
FT NON_TER 360
SQ SEQUENCE 360 AA; 40170 MW; 224E3CD345BD2F7A CRC64;
Query Match 19.3%; Score 1625.5; DB 5; Length 360;
Best Local Similarity 90.0%; Pred. No. 5.2e-44;
Matches 325; Conservative 12; Mismatches 19; Indels 5; Gaps 3;
Qy 1 MKIIFFLCSFLFFINTOCVTHESYOELVKKLEALEDVLTGYSLFQKEKMWLNEGTSGT 60
Db 1 MKIIFFLCSFLFFINTOCVTHESYOELVKKLEALEDVLTGYSLFQKEKMWLNEGTSGT 60
Qy 61 AVTTSTPGSKGVSASGGSGGVA---SGGSVASGSGVASGSGGSGNSRRTNPDSNS 117
Db 61 AVTTSTPGSKGVSVSGSGGSGGVA---SGGSVASGSGVASGSGGSGNSRRTNPDSNS 120
Qy 118 SDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCDNIHGFYKYLIDGYEEINEL 177
Db 121 SDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCDNIHGFYKYLIDGYEEINEL 180
Qy 178 LYKLNFFDRLAKLNDVCANDYCOIPENLKITRANELDVLKLVFGYRKPLDNIKDNVYK 237
Db 181 LYKLNFFDRLAKLNDVCANDYCOIPENLKITRANELDVLKLVFGYRKPLDNIKDNVYK 240
Qy 238 MEDYIKKNNKTITANINELIEESKKTIDKNKATKEEKKLYQAOYDLSIYNKQLEEAHN 297
Db 241 MEDYIKKNNKTITANINELIEESKKTIDQKNADNEEGKKLYQAOYDLSIYNKQLEEAHN 300
Qy 298 LISVLEKRIDTLKKNENIKELLDKINEIK-NPPPPANSNGTPTNLDDKNKIEEHEKEIKE 356
Db 301 LISVLEKRIDTLKKNENIKELLDKIDKIDAEKLTGSKPNTLLDKNKIEEHEKEIKE 359
Qy 357 I 357
```

```
Db 360 I 360
RESULT 45
Q9TZU9 PRELIMINARY; PRT; 363 AA.
AC Q9TZU9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFA11;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061139; AAC69738.1; -.
FT NON_TER 363
SQ SEQUENCE 363 AA; 40398 MW; C25F38CB636941B7 CRC64;
Query Match 19.3%; Score 1625.5; DB 5; Length 363;
Best Local Similarity 89.0%; Pred. No. 5.3e-44;
Matches 323; Conservative 15; Mismatches 18; Indels 7; Gaps 3;
Qy 1 MKIIFFLCSFLFFINTOCVTHESYOELVKKLEALEDVLTGYSLFQKEKMWLNEGTSGT 60
Db 1 MKIIFFLCSFLFFINTOCVTHESYOELVKKLEALEDVLTGYSLFQKEKMWLNEGTSGT 60
Qy 61 AVTTSTPGSKGVSASGGSGGVA---SGGSVASGSGVASGSGGSGNSRRTNPNS 114
Db 61 AVTTSTPGSKGSGSVASGGSGGSGVASGSGGSGVASGSGGSGNSRRTNPNS 120
Qy 115 DNSSDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCDNIHGFYKYLIDGYEEI 174
Db 121 DNSSDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCDNIHGFYKYLIDGYEEI 180
Qy 175 NELLYKLNFFDRLAKLNDVCANDYCOIPENLKITRANELDVLKLVFGYRKPLDNIKDN 234
Db 181 NELLYKLNFFDRLAKLNDVCANDYCOIPENLKITRANELDVLKLVFGYRKPLDNIKDN 240
Qy 235 VGMEDYIKKNNKTITANINELIEESKKTIDKNKATKEEKKLYQAOYDLSIYNKQLEE 294
Db 241 VGMEDYIKKNNKTITANINELIEESKKTIDQKNADNEEGKKLYQAOYDLSIYNKQLE 300
Qy 295 AHNLSVLEKRIDTLKKNENIKELLDKINEIK-NPPPPANSNGTPTNLDDKNKIEEHEKE 353
Db 301 AHNLSVLEKRIDTLKKNENIKELLDKIDKIDAEKLTGSKPNTLLDKNKIEEHEKE 360
Qy 354 IKE 356
Db 361 IKE 363
RESULT 46
Q9TZU1 PRELIMINARY; PRT; 344 AA.
AC Q9TZU1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```


OX	NCBI_TaxID=5833;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=HN6.27;
RX	MEDLINE=20106724; PubMed=10643908;
RA	Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA	Pluschke G.;
RT	"Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT	falciparum in clinical isolates from the Kilombero District,
RT	Tanzania.";
RL	Acta Trop. 74:51-61(2000).
DR	EMBL; AF061147; AAC69746.1; -.
FT	NON_TER 344 344
SQ	SEQUENCE 344 AA; 38767 MW; 1225C0E1683007DF CRC64;

Query Match	19.3%; Score 1622; DB 5; Length 344;
Best Local Similarity	89.6%; Pred. No. 6.4e-44;
Matches 319; Conservative	11; Mismatches 14; Indels 12; Gaps 2;

Oy	1 MKIIFFLCSPFFIINTQCVTTHESYQELVKKLEALEDAVL7GYSLFQEKVMVNEGTSCT 60
Dd	
Db	1 MKIIFFLCSPFFIINTQCVTTHESYQELVKKLEALEDAVL7GYSLFQEKVMVNEGTSCT 60
Oy	61 AVTTSTPGSGSVASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTPNSDSSDS 120
Dd	
Db	61 AVTTSTPGSGSVTSGGGGSVA-----SVA---SVASGGSGSNRRTPNSDSSDS 108
Oy	121 DAKSYADLKHRVNVLTKIKELYPOLFDLTNNHMLTCDNIHGPKYLIDGYEEINELLYK 180
Dd	
Db	109 DAKSYADLKHRVQNVLTIKELKPPELFDLTNNHMLTCDNIHGPKYLIDGYEEINELLYK 168
Oy	181 LNFYEDLLRAKLDVCANDYCOIPFNLIKIRANELDLVKLVFGYRKPLDNIDKNVGKMED 240
Dd	
Db	169 LNFYEDLLRAKLDACANSYCOIPFNLIKIRANELDLVKLVFGYRKPLDNIDKNVGKMED 228
Oy	241 YIKNKKTITENINELIEBSKTDIOKNKATVEBKKLQAYDLISITYNKOLEAHNLIS 300
Dd	
Db	229 YIKNKKTITANINELIEGSEKTDIQNKADNEEGKKLYQAYNLFIYNKQLQEAHNLIS 288
Oy	301 VLEKRIDTLKKNENTKELLDKINETKNPPPSANGTPTNTLLDKNKKIEBEKEIKE 356
Dd	
Db	289 VLEKRIDLAKKNENIKLLGLKINETKNPPPSANGTPTNTLLDKNKKIEBEKEIKE 344

RESULT 47	
Q9TZV3	PRELIMINARY; PRT; 362 AA.
ID Q9TZV3	PRELIMINARY;
AC Q9TZV3;	
DT 01-MAY-2000 (TrEMBLrel. 13, Created)	
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE Merozoite surface protein 1 (Fragment).	
GN MSPI.	
OS Plasmodium falciparum.	
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX NCBI_TaxID=5833;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=IFA9.16;	
RC MEDLINE=20106724; PubMed=10643908;	
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,	
RA Pluschke G.;	
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium	
RT falciparum in clinical isolates from the Kilombero District,	
RT Tanzania.";	
RL Acta Trop. 74:51-61(2000).	
DR EMBL; AF061135; AAC69734.1; -.	
FT NON_TER 362 362	
SQ SEQUENCE 362 AA; 41401 MW; A31290A0C41AC600 CRC64;	

Query Match	19.2%; Score 1616.5; DB 5; Length 362;
Best Local Similarity	86.8%; Pred. No. 1.e-43;
Matches 322; Conservative	11; Mismatches 29; Indels 9; Gaps 2;


```
Db 181 NELLVYKLFYFOLLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRKPLDNKDN 240
QY 235 VGMEDYIKNNKTTIENIELIEESKTTIDKNKNATKEEKKLYQAQYDLSIYNKQLEE 294
Db 241 VGMEDYIKNNKTTIANINELIEGSKTTIDQKNADNEEGKKLYQAQYDLSIYNKQLEE 300
QY 295 AHNLSVLEKRDITLKKNNENIKELDKINEIK-NPPANSNGTPTNTLLD-KNKKTEEHEK 352
Db 301 AHNLSVLEKRDITLKKNNENIKKLEDDIKTDAEKLTTGSKPNPLPENKKEVEGHEE 360
QY 353 EIKEI 357
Db 361 KIKEI 365

RESULT 49
Q9TZU4
ID Q9TZU4 PRELIMINARY; PRT; 351 AA.
AC Q9TZU4;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HN3;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL: AF061144; AAC69744.1; -.
FT NON_TER 351
SQ SEQUENCE 351 AA; 39465 MW; 06481719E004278D CRC64;

Query Match 18.8%; Score 1585.5; DB 5; Length 351;
Best Local Similarity 88.2%; Pred. No. 9e-43;
Matches 315; Conservative 16; Mismatches 19; Indels 7; Gaps 3;

QY 1 MKIIFLCSFLFFIINTQCVTTHESYQELVKLEALEDAVLTCYSLFQKEKMWLNESGTS 60
Db 1 MKIIFLCSFLFFIINTQCVTTHESYQELVKLEALEDAVLTCYSLFQKEKMWLNESGTS 60
QY 61 AVTTTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSG 120
Db 61 AVTTTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSG 114
QY 121 DAKSYADLKHRYNLLTIKELYPQLFDLTNHMLTLCNDNIHGFYKYLIDGYEEINELLYK 180
Db 115 DAKSYADLKHRYNLLTIKELYPQLFDLTNHMLTLCNDNIHGFYKYLIDGYEEINELLYK 174
QY 181 LNFYDILLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRKPLDNKDNVGMED 240
Db 175 LNFYDILLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRKPLDNKDNVGMED 234
QY 241 YIKNNKTTIENIELIEESKTTIDKNKNATKEEKKLYQAQYDLSIYNKQLEEANLIS 300
Db 235 YIKNNKTTIANINELIEGSKTTIDQKNADNEEGKKLYQAQYDLSIYNKQLEEANLIS 294
QY 301 VLEKRDITLKKNNENIKELDKINEIK-NPPANSNGTPTNTLLDKNKKIEEHEKEIKE 356
Db 295 VLEKRDITLKKNNENIKKLEDDIKTDAEKLTTGSKPNPLPENKKEVEHEKEIKE 351

RESULT 51
Q9TZV0
ID Q9TZV0 PRELIMINARY; PRT; 360 AA.
AC Q9TZV0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IFA10;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
```

```
ID Q9TZU3 PRELIMINARY; PRT; 351 AA.
AC Q9TZU3;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HN3;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL: AF061145; AAC69744.1; -.
FT NON_TER 351
SQ SEQUENCE 351 AA; 39465 MW; 06481719E004278D CRC64;

Query Match 18.8%; Score 1585.5; DB 5; Length 351;
Best Local Similarity 88.2%; Pred. No. 9e-43;
Matches 315; Conservative 16; Mismatches 19; Indels 7; Gaps 3;

QY 1 MKIIFLCSFLFFIINTQCVTTHESYQELVKLEALEDAVLTCYSLFQKEKMWLNESGTS 60
Db 1 MKIIFLCSFLFFIINTQCVTTHESYQELVKLEALEDAVLTCYSLFQKEKMWLNESGTS 60
QY 61 AVTTTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSG 120
Db 61 AVTTTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSG 114
QY 121 DAKSYADLKHRYNLLTIKELYPQLFDLTNHMLTLCNDNIHGFYKYLIDGYEEINELLYK 180
Db 115 DAKSYADLKHRYNLLTIKELYPQLFDLTNHMLTLCNDNIHGFYKYLIDGYEEINELLYK 174
QY 181 LNFYDILLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRKPLDNKDNVGMED 240
Db 175 LNFYDILLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRKPLDNKDNVGMED 234
QY 241 YIKNNKTTIENIELIEESKTTIDKNKNATKEEKKLYQAQYDLSIYNKQLEEANLIS 300
Db 235 YIKNNKTTIANINELIEGSKTTIDQKNADNEEGKKLYQAQYDLSIYNKQLEEANLIS 294
QY 301 VLEKRDITLKKNNENIKELDKINEIK-NPPANSNGTPTNTLLDKNKKIEEHEKEIKE 356
Db 295 VLEKRDITLKKNNENIKKLEDDIKTDAEKLTTGSKPNPLPENKKEVEHEKEIKE 351

RESULT 51
Q9TZV0
ID Q9TZV0 PRELIMINARY; PRT; 360 AA.
AC Q9TZV0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IFA10;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
```


RT	falciparum in clinical isolates from the Kilombero District, Tanzania. ² ;									
RL	Acta Trop.	74:51-61(2000).								
DR	EMBL:	AF061142; AAC69741.1; -.								
FT	NON_TER	352	352							
SEQ	SEQUENCE	352 AA; 39477 MW; 1C62B4ED3025175B	CRC64;							
Query Match				18.3%;	Score 1544;	DB 5;	Length 352;			
Best Local Similarity				86.0%;	Pred. No. 1.9e-41;					
Matches 308; Conservative				19;	Mismatches 23;	Indels	8;	Gaps		
QY	1	MKIIFFLCSELPFIINTQCVTHESYQELVKKLEALEDAVL	TGYSLFQKMKVINEGTS	60						
DB	1	MKIIFFLCSELPFIINTQCVTHESYQELVKKLEALEDAVL	TGYSLFQKMKVINEGTS	60						
QY	61	AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGV	SGSGSRRTNPSD	120						
DB	61	AVTTSTPGSGVSTSGSGGSA--SVASG--	SGSGSRRTNPSD	114						
QY	121	DAKSYADLKHVRNRYLTIKELYPQLFDLTNHLMTLCD	NIHGFYLDGYEINELLYK	180						
DB	115	DAKSYADLKHVRQNYLFTIKELYPPELFDLTNHLMTLCD	NIHGFYLDGYEINELLYK	174						
QY	181	LNIFYDLRAKLNDVCANDYCOIPFNLIKIRANELDVL	KLKLVFGYRPLDNKIVGCKMED	240						
DB	175	LNIFYDLRAKLNDACANSYCOIPFNLIKIRANELDVL	KLKLVFGYRPLDNKIVGCKMED	234						
QY	241	YIKNNKTTIENINELTBESKTTIDKNKNATYKKEBK	KLYQAOYDLSYNNKQLEAHNLIS	300						
DB	235	YIKNNKTTIANINELTBESKTTIDKNKNADNEGKK	LYQAOYDLSYNNKQLEAHNLIS	294						
QY	301	VLEKRIDTLKKNENIKELLDKINIEIK-NPPANS	GNTPNTLLD-KNKKIEEHEKEIKE	356						
DB	295	VLEKRIDTLKKNENIKELLDKIKITDAEKLTTGSK	PNPLPENKKKEVGHEKEIKE	352						
RESULT 56										
ID	Q9TZV4	PRELIMINARY;	PRT;	347 AA.						
AC	Q9TZV4;									
DT	01-MAY-2000	(Tremblrel. 13, Created)								
DT	01-MAY-2000	(Tremblrel. 13, Last sequence update)								
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)								
DE	Merozoite surface protein 1 (Fragment).									
GN	MSP1.									
OS	Plasmodium falciparum.									
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.									
OX	NCBI_TaxID=5833;									
RN	[1]									
RC	SEQUENCE FROM N.A.									
RP	STRAIN=IPAS.5;									
RX	MEDLINE=20106724; PubMed=10643908;									
RA	Jiang G., daubenberger C., Huber W., Matile H., Tanner M., Pluschke G.;									
RT	*Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania. ² ;									
RL	Acta Trop.	74:51-61(2000).								
DR	EMBL:	AF061134; AAC69733.1; -.								
FT	NON_TER	347	347							
SEQ	SEQUENCE	347 AA; 39693 MW; 75DCA3237A9B7E33	CRC64;							
Query Match				18.3%;	Score 1542.5;	DB 5;	Length 347;			
Best Local Similarity				86.2%;	Pred. No. 1.9e-41;					
Matches 307; Conservative				11;	Mismatches 29;	Indels	9;	Gaps		
QY	1	MKIIFFLCSELPFIINTQCVTHESYQELVKKLEALEDAVL	TGYSLFQKMKVINEGTS	60						
DB	1	MKIIFFLCSELPFIINTQCVTHESYQELVKKLEALEDAVL	TGYSLFQKMKVINEEIT	60						
QY	61	AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGV	SGSGSRRTNPSD	120						
DB	61	KGASAOGTSGT--SGTSGSPGSGTSPSSRSRTLP	SRNTSSCA-----PPADASDS	111						

RESULT 58		PRELIMINARY;		PRT;	344 AA.
ID	Q9T2T9				
AC	Q9T2T9				
DT	01-MAY-2000 (TREMBlrel. 13, Created)				
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)				
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)				
DE	Merozoite surface protein 1 (fragment).				
GN	MSPI.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5833;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=IFA7;				
RX	MEDLINE=20106724; PubMed=10643908;				
RA	Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,				
RA	Pluschke G.;				
RT	"Sequence diversity of the merozoite surface protein 1 of Plasmodium				
RT	falciparum in clinical isolates from the Kilombero District,				
RT	Tanzania.";				
RL	Acta Trop. 74:51-61(2000).				
DR	EMBL; AF061149; AAC69748.1; -.				
FT	NON_TER 344 344				
SQ	SEQUENCE 344 AA; 39227 MW; 9E79A793662667DB CRC64;				
Query Match 17.78; Score 1488; DB 5; Length 344;					
Best Local Similarity 82.6%; Pred. No. 9.7e-40;					
Matches 294; Conservative 15; Mismatches 35; Indels 12; Gaps					
Qy	1	MKIIFFLCSLFFFIINTOCVTHESYQELVKKLEALEDAVLGYSILFQKRMVLNEGTS	60		
Db	1	MKIIFFLCSLFFFIINTOCVTHESYQELVKKLEALEDAVLGYSILFQKRMVLKDGANTQ	60		
Qy	61	AVTSTPGSKSVASGGSGSVASGGSVASGGSVASGGSGNSRRTPNSDSSDS	12		
Db	61	VVAKPADAVSTQSAKNPPGATVPS-----GTASTKGAIRSPD-----ANPSDDSSDS	10		
Qy	121	DAKSYADLKHVRNYLLTIKELKYPQLPFDLTNNHMLTLCDNIGHGFKYLDGYEENELLYK	18		
Db	109	DAKSYADLKHVRQNYLFTIKELKYPQLPFDLTNNHMLTLCDNIGHGFKYLDGYEENELLYK	16		
Qy	181	LNFTFDLLRAKLDNCANDYQIQIFNLKIRANELDVLKLVGFRKKPLDNTKDNVQKMD	24		
Db	169	LNFTFDLLRAKLDNCANDYQIQIFNLKIRANELDVLKLVGFRKKPLDNTKDNVQKMD	22		
Qy	241	YIKNKKTIENIELIESKTTIDKNKATKEEKKLYQAYDLSYNKOLEEAHNLS	30		
Db	229	YIKNKKTIENIELIESKTTIDQNKNADNEEGKKLYQAYDLSYNKOLEEAHNLS	28		
Qy	301	VLEKRIDFLKKNENIKELDKINEIKNPPPSNGTPTNLLDKNKKIEEHEKEIKE	356		
Db	289	VLEKRIDFLKKNENIKELDKINEIKNPPPSNGTPTNLLDKNKKIEEHEKEIKE	344		
RESULT 59					
ID	Q25760				
AC	Q25760				
DT	01-NOV-1996 (TREMBlrel. 01, Created)				
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)				
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)				
DE	Merozoite surface protein 1 (fragment).				
GN	MSPI.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5833;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91304517; PubMed=1852173;				
RA	Kerr P.J., Ranford-Cartwright L.C., Walliker D.;				
RT	"Proof of intragenic recombination in Plasmodium falciparum.";				
RL	Mol. Biochem. Parasitol. 46:185-187(1991).				

```
RN [2]
RP -SEQUENCE FROM N.A.
RX MEDLINE=95107347; PubMed=7808474;
RA Kerr P.J., Ranford-Cartwright L.C., Walliker D.;
RT "Proof of intragenic recombination in Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 66:241-248(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Ranford-Cartwright L.C.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X52963; CAA37136.1; -
FT NON_TER 1
FT NON_TER 363
SQ SEQUENCE 363 AA; 40336 MW; A86C08A93DE5AF9B CRC64;

Query Match      17.4%; Score 1467.5; DB 5; Length 363;
Best Local Similarity 81.6%; Pred. No. 4.4e-39;
Matches 298; Conservative 10; Mismatches 38; Indels 19; Gaps 4;

QY 21 THESYQELVKKLEALEDVLTGYSLFQKEKMWLNE---GTSCTAVTTSTPGSKGVSAGG 77
Db 1 THESYQELVKKLEALEDVLTGYSLFQKEKMWLNEEITTKGASQAQSGA--ASQA 58

QY 78 SGGSVASGGSVASGGSVASGGSGN--SRRTN-----PSDNSSDSDAK 123
Db 59 SGASAQSGASAQSGASAQSGTSGPSGSPSRNTLPRSNLTSSGASPPADASDSA 118

QY 124 SYADLKHRVRYLLTIKELKYPQLFDLTNHHMLTLCDNHGHFKYLDGYEINELLYKLN 183
Db 119 SYADLKHRVRYLLTIKELKYPQLFDLTNHHMLTLCDNHGHFKYLDGYEINELLYKLN 178

QY 184 YFDLLRAKLVNDYANDYCOIPENLIRANELDVLLKLVFGYRKPLDNKDNVGMEDYIK 243
Db 179 YFDLLRAKLVNDYANDYCOIPENLIRANELDVLLKLVFGYRKPLDNKDNVGMEDYIK 238

QY 244 KKKTTIENIELIESKKTIDKNKNATKEEEKKKLYQAOYDLSYNNKQLEEAHNLISVLE 303
Db 239 KKKTTIANIELIEGSKKTIDQKNADNEEGKKLYQAOYDLSYNNKQLEEAHNLISVLE 298

QY 304 KRIDLKKNENIKELLDKINEIKNPPANGNTPTNLLDKNKKIEEHEKEIKEIAKTIRF 363
Db 299 KRIDLKKNENIKLLDKINEIKNPPANGNTPTNLLDKNKKIEEHEKEIKEIAKTIRF 358

QY 364 NIDSL 368
Db 359 NIDSL 363
```

Search completed: March 31, 2003, 07:25:26
Job time : 104 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 04:33:12 : Search time 25 Seconds
(without alignments)
2719.188 Million cell updates/sec

Title: US-09-269-874A-3

Perfect score: 8424

Sequence: 1 MKIIFLCSEFLFIINTQCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	8424	100.0	1639	1	MSPI_PLAFW	P04933 plasmodium
2	8144.5	96.7	1630	1	MSPI_PLAFK	P04932 plasmodium
3	5136.5	61.0	1726	1	MSPI_PLAFEC	P04934 plasmodium
4	5127.5	60.9	1726	1	MSPI_PLAFEP	P50495 plasmodium
5	5069	60.2	1701	1	MSPI_PLAFM	P08569 plasmodium
6	5065	60.1	1701	1	MSPI_PLAFM	P13819 plasmodium
7	4854.5	57.6	1682	1	MSPI_PLAF3	P19598 plasmodium
8	2548.5	30.3	1772	1	MSPI_PLAFYO	P13828 plasmodium
9	1187	14.1	233	1	MSPI_PLAFD	P13827 plasmodium
10	1096.5	13.0	281	1	MSPI_PLAFN	P13820 plasmodium
11	501.5	6.0	1957	1	YD86_SCHPO	Q10411 schizosacch
12	486	5.8	1790	1	USO1_YEAST	P25386 saccharomyc
13	477	5.7	2869	1	RBP1_PLAVB	Q00798 plasmodium
14	474.5	5.6	1875	1	MLP1_YEAST	Q02455 saccharomyc
15	457.5	5.4	1251	1	RBP2_PLAVB	Q00799 plasmodium
16	443.5	5.3	2663	1	CENE_HUMAN	Q02224 homo sapien
17	440	5.2	1805	1	HMW2_MYCGE	P47460 mycoplasma
18	424	5.0	2230	1	GOG4_HUMAN	Q13439 homo sapien
19	423	5.0	1679	1	YIO9_YEAST	P40457 saccharomyc
20	421.5	5.0	1928	1	MSY1_YEAST	P08964 saccharomyc
21	405	4.8	2022	1	ANTI_ONCVO	P21249 onchocerca
22	404	4.8	2116	1	MSY2_DICDI	P08799 dictyosteli
23	399.5	4.7	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
24	393	4.7	1312	1	RA50_YEAST	P12753 saccharomyc
25	386.5	4.6	2748	1	NUH1_YEAST	Q00402 saccharomyc
26	378	4.5	1818	1	HMW2_MYCPN	P75471 mycoplasma
27	372	4.4	1005	1	RA50_METJA	Q58718 methanococc
28	365	4.3	3911	1	AKA9_HUMAN	Q99996 h a-kinase
29	361	4.3	3660	1	DMD_CHICK	P11533 gallus gall
30	358.5	4.3	2349	1	TFR_HUMAN	P12270 homo sapien
31	355.5	4.2	3210	1	CENF_HUMAN	P49454 homo sapien
32	353	4.2	1169	1	SNC_METJA	Q59037 methanococc
33	352.5	4.2	1163	1	SBCC_CLOAB	Q97fk1 clostridium

P02564 rattus norv
P12883 homo sapien
P34367 caenorhabdi
P13539 mesocricetu
P02565 gallus gall
P32908 saccharomyc
O67124 aquifex aeo
P13540 mesocricetu
P09975 marchantia
P13533 homo sapien
Q02566 mus musculu
P10587 gallus gall
O51578 borrelia bu
P02563 rattus norv
Q03061 homo sapien
O15078 homo sapien
P10622 homo sapien
P11532 homo sapien
P15924 homo sapien
P05661 drosophila
P79293 sus scrofa
Q15431 homo sapien
Q97wh0 sulfolobus
P24733 aequiptecten
P13538 gallus gall
Q01550 xenopus lae
P54697 dictyosteli
Q03410 rattus norv
Q9ukx3 homo sapien
Q9tu23 bos taurus
Q96yr5 sulfolobus
P12882 homo sapien
P32380 saccharomyc
P29616 gallus gall
Q28641 cryotolagus
Q92351 schizosacch
P27625 plasmodium
P38989 saccharomyc
P35580 homo sapien
Q99323 drosophila
Q62209 mus musculu
Q33600 sulfolobus
P58301 pyrococcus
P13535 homo sapien
Q14157 schizosacch
O94623 homo sapien
Q03661 saccharomyc
Q97592 canis famli
Q971t0 rattus norv
P12847 rattus norv
Q27991 bos taurus
Q04956 plasmodium
P02562 cryotolagus
Q99104 mus musculu
Q60563 mesocricetu
Q24702 dictyocaulu
Q49419 mycoplasma
P97929 mus musculu
P20929 homo sapien
P27895 saccharomyc
P38198 saccharomyc
Q9cfz0 lactococcus
P11055 homo sapien
Q90988 gallus gall
Q90339 cyprinus ca
P16154 clostridium
Q08372 plasmodium
Q042184 gallus gall
P11531 mus musculu
Q27171 paramecium
P50532 xenopus lae
P46939 homo sapien
Q11102 caenorhabdi


```
QY 1 MKIIFFLCSPFFIIINTQCVTTHESQELVKLEALEDAVLGTGYSLFQKEMVINEGTSOT 60
Db 1 MKIIFFLCSPFFIIINTQCVTTHESQELVKLEALEDAVLGTGYSLFQKEMVINEGTSOT 60
QY 61 AVTTSTPGSKSVASGGSGSVASGGSVASGGSVASGGSVASGGSGNSRRTPNSDSSDS 120
Db 61 AVTTSTPGSKSVASGGSGSVASGGSVASGGSVASGGSVASGGSGNSRRTPNSDSSDS 120
QY 121 DAKSVADLKHVRNRYLLTIKELKYPOLFDLTNHHMLTLCNTIHFYKYLIDGYEIEINELLYK 180
Db 121 DAKSVADLKHVRNRYLLTIKELKYPOLFDLTNHHMLTLCNTIHFYKYLIDGYEIEINELLYK 180
QY 181 LNFYFDLLRAKLVNDVANDYCOIPNLKIRANELDLVKKLVFGYRKPLDNTKDNVGRMED 240
Db 181 LNFYFDLLRAKLVNDVANDYCOIPNLKIRANELDLVKKLVFGYRKPLDNTKDNVGRMED 240
QY 241 YIKKNNKKTENINELIEESKTTIDKNKNATKEERKKLYOAOYDLISYINKOLEEAHNLS 300
Db 241 YIKKNNKKTENINELIEESKTTIDKNKNATKEERKKLYOAOYDLISYINKOLEEAHNLS 300
QY 301 VLEKRIDTLKKNENIKELDKINEIKNPPANGSGTPTNTLLDKNKKIEBEHEKEIKEIAKT 360
Db 301 VLEKRIDTLKKNENIKELDKINEIKNPPANGSGTPTNTLLDKNKKIEBEHEKEIKEIAKT 360
QY 361 IKFNIDSFTDPLELEYLREKNNKIDISAKVETKESTEPNEYPNGVTYPLSYNDINNAL 420
Db 361 IKFNIDSFTDPLELEYLREKNNKIDISAKVETKESTEPNEYPNGVTYPLSYNDINNAL 420
QY 421 NELNSFGDLINPFDYTKPSKNIYTDNERKKFPIEIKIEKIEKKKIESDKKSYEDRSKS 480
Db 421 NELNSFGDLINPFDYTKPSKNIYTDNERKKFPIEIKIEKIEKKKIESDKKSYEDRSKS 480
QY 481 LNDITKEYEKLNEIYDSKFNNNIDLTFNFKMMGRYKYVEKLTPHHNTFASYENSKNKL 540
Db 481 LNDITKEYEKLNEIYDSKFNNNIDLTFNFKMMGRYKYVEKLTPHHNTFASYENSKNKL 540
QY 541 EKLTKALYKMEYSLURNIWEKELYKYNLISKIENEIETLVNTKDEEQLFKKITKD 600
Db 541 EKLTKALYKMEYSLURNIWEKELYKYNLISKIENEIETLVNTKDEEQLFKKITKD 600
QY 601 ENKPDKEILEYSDIVKVOQVLLMANKIDELKKTOLILKNVELKHNHVPNSYKOEKOE 660
Db 601 ENKPDKEILEYSDIVKVOQVLLMANKIDELKKTOLILKNVELKHNHVPNSYKOEKOE 660
QY 661 PYLIVLKKEDIKLVFMPKVESLINEEKKNIKTEGQSDNSEPSTEGBITQOATTKPGQ 720
Db 661 PYLIVLKKEDIKLVFMPKVESLINEEKKNIKTEGQSDNSEPSTEGBITQOATTKPGQ 720
QY 721 AGSALEGDSVQAQOEQAQOPPPVPPVPEAKAQVTPPAVNNKNTENVSKLDYLEKLYE 780
Db 721 AGSALEGDSVQAQOEQAQOPPPVPPVPEAKAQVTPPAVNNKNTENVSKLDYLEKLYE 780
QY 781 FLNTSYICHKYTLVSHSTWNEKILKOYKITDEESKLSKCDPLDLFLNQNINIPVMSMF 840
Db 781 FLNTSYICHKYTLVSHSTWNEKILKOYKITDEESKLSKCDPLDLFLNQNINIPVMSMF 840
QY 841 DSLNNSLSOLFMEIYEKEMVCNLYKLKNDKIKNLEEAQVSTSVKTLSSSSMQPLSLT 900
Db 841 DSLNNSLSOLFMEIYEKEMVCNLYKLKNDKIKNLEEAQVSTSVKTLSSSSMQPLSLT 900
QY 901 PQDRPEVSANDTSHSTNLNLSKLIFENILSLGKNKIYQELIGKSSSENFYEKILKQSD 960
Db 901 PQDRPEVSANDTSHSTNLNLSKLIFENILSLGKNKIYQELIGKSSSENFYEKILKQSD 960
QY 961 TFYNESFTNFVKSADDTINSLNDESKRKKLEEDINKLAKTTLQSLFDLYNKYKLERLFD 1020
Db 961 TFYNESFTNFVKSADDTINSLNDESKRKKLEEDINKLAKTTLQSLFDLYNKYKLERLFD 1020
QY 1021 KKKTVGKYMOKIKLTLKLEQLESKLSLNPNPKHVLFQNSVFFNKKKEAEIAETENTLEN 1080
Db 1021 KKKTVGKYMOKIKLTLKLEQLESKLSLNPNPKHVLFQNSVFFNKKKEAEIAETENTLEN 1080
QY 1081 TKILLKHYKGLVKYNGESSPLKTLSESIQEDNYASLENFKVLKLEGLKDNLNLEK 1140
```

```
Db 1081 TKILLKHYKGLVKYNGESSPLKTLSESIQEDNYASLENFKVLKLEGLKDNLNLEK 1140
QY 1141 KKLVSLSGLHLIAELAEVINKNKYTGNSPSENNTDVNNALLESYKKFLPEGTDVATVVS 1200
Db 1141 KKLVSLSGLHLIAELAEVINKNKYTGNSPSENNTDVNNALLESYKKFLPEGTDVATVVS 1200
QY 1201 ESGDFTLQSQPKPASTHVGAESNTITTSQNVDEVDVVIIVPIFGESEEDYDDLQGVV 1260
Db 1201 ESGDFTLQSQPKPASTHVGAESNTITTSQNVDEVDVVIIVPIFGESEEDYDDLQGVV 1260
QY 1261 TGEAVTPSVIDNLSKIENEYEVLYKPLAGVYRSLKOLENNVMTFNVNVDILNSRPN 1320
Db 1261 TGEAVTPSVIDNLSKIENEYEVLYKPLAGVYRSLKOLENNVMTFNVNVDILNSRPN 1320
QY 1321 KRENEKNVLESOLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSYNYIKDSITDIDNAN 1380
Db 1321 KRENEKNVLESOLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSYNYIKDSITDIDNAN 1380
QY 1381 DVLGYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFNNIETLYKTVNDKIDLFVHLE 1440
Db 1381 DVLGYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFNNIETLYKTVNDKIDLFVHLE 1440
QY 1441 AKVLNYYTEKSNVEVKIKELNYLKTIDQKLADFKNNNPVGADLSTDYNNHNNLTKFUS 1500
Db 1441 AKVLNYYTEKSNVEVKIKELNYLKTIDQKLADFKNNNPVGADLSTDYNNHNNLTKFUS 1500
QY 1501 TGMVFNENAKTVLSNLDGNOGMLNISOHCQVKKOCPOKSCFRHLDERECKCLLNK 1560
Db 1501 TGMVFNENAKTVLSNLDGNOGMLNISOHCQVKKOCPOKSCFRHLDERECKCLLNK 1560
QY 1561 QEGDKVCENPNPTCENNGGCDADAKTEEDSGSNGKKTICETCKPDSYPLFDGIFCSSS 1620
Db 1561 QEGDKVCENPNPTCENNGGCDADAKTEEDSGSNGKKTICETCKPDSYPLFDGIFCSSS 1620
QY 1621 NPLGISFLLILMLILYSFI 1639
Db 1621 NPLGISFLLILMLILYSFI 1639
RESULT 2
MSPL_PLAFK STANDARD: PRT: 1630 AA.
ID MSPL_PLAFK AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate KI / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X03371; CAA37070.1; -

DR PIR; A25120; SAZQK1.

DR InterPro; IPR000561; EGF-like.

DR Pfam; PF00008; EGF; 1.

KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;

KW Transmembrane; GPI-anchor.

FT SIGNAL 1 19

FT CHAIN 20 1630 POTENTIAL.

FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.

FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.

FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;

Query Match 96.7%; Score 8144.5; DB 1; Length 1630;
Best Local Similarity 97.1%; Pred. No. 3.6e-258;
Matches 1591; Conservative 11; Mismatches 28; Indels 9; Gaps 2;

QY 1 MKTIIFLCFLPIINTQCVTHESYQELVKKLEALDAVLTYGSLFQKEXKMWLNCTSGT 60
DB 1 MKTIIFLCFLPIINTQCVTHESYQELVKKLEALDAVLTYGSLFQKEXKMWLNCTSGT 60
QY 61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSRRTPNSDSSDS 120
DB 61 KGASQSGTSGT--SQTSGPSGSGTSPSRNTLPSRNTSSGAS-----PPADASD 111
QY 121 DAKSYADLKHVRNYLLTIKELYPQLFDLTNHLMTLCNDIHGFKYLDIGYEEINELLYK 180
DB 112 DAKSYADLKHVRNYLLTIKELYPQLFDLTNHLMTLCNDIHGFKYLDIGYEEINELLYK 171
QY 181 LNFYFDLLRAKLDVNCANDYQIIPFNLIKTRANELDVLLKLVFGYRKPDLNDKGVKMD 240
DB 172 LNFYFDLLRAKLDVNCANDYQIIPFNLIKTRANELDVLLKLVFGYRKPDLNDKGVKMD 231
QY 241 YIKKKNKTTENINELLEESKKTIDKKNKATKEBEKKLYQAOYDLSIYNKQLPEAHNLIS 300
DB 232 YIKKKNKTTENINELLEESKKTIDKKNKATKEBEKKLYQAOYDLSIYNKQLPEAHNLIS 291
QY 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNTLLDKNKKIEEHEKEIKEIAKT 360
DB 292 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNTLLDKNKKIEEHEKEIKEIAKT 351
QY 361 IKFNIDSLTDPLEYILREKNKNIDISAKVETKESTPEPNEGVTYPLSYNDINNAL 420
DB 352 IKFNIDSLTDPLEYILREKNKNIDISAKVETKESTPEPNEGVTYPLSYNDINNAL 411
QY 421 NELNSFGDLINPDYTKPSKNITYONERKKFINETKEKIKTEKKIESDKKSYEDRSKS 480
DB 412 NELNSFGDLINPDYTKPSKNITYONERKKFINETKEKIKTEKKIESDKKSYEDRSKS 471
QY 481 LNDITKEYEKLLEIYDSKFNNDIDITNFKMMGKRYSYKVEKLTHHNTFASYSKHNH 540
DB 472 LNDITKEYEKLLEIYDSKFNNDIDITNFKMMGKRYSYKVEKLTHHNTFASYSKHNH 531
QY 541 EKLTKALKYMEDYSLRNIVVEKELKYKKNLI SKIENIEITLVENIKKDEQLFEKKITKD 600
DB 541 EKLTKALKYMEDYSLRNIVVEKELKYKKNLI SKIENIEITLVENIKKDEQLFEKKITKD 600

DB 532 EKLTKALKYMEDYSLRNIVVEKELKYKKNLI SKIENIEITLVENIKKDEQLFEKKITKD 591
QY 601 ENKPDEKILEVSDIVKQVQKVLMMNKIDELKKTQILTKNVELKHNHIVPNSYKQENKQE 660
DB 592 ENKPDEKILEVSDIVKQVQKVLMMNKIDELKKTQILTKNVELKHNHIVPNSYKQENKQE 651
QY 661 PYYLIVLKKKIDKLVFMPKVESLINEEKNIKTEGSDNSSEPTGEITGQATTPGQQ 720
DB 652 PYYLIVLKKKIDKLVFMPKVESLINEEKNIKTEGSDNSSEPTGEITGQATTPGQQ 711
QY 721 AGSALEGDSVQAQAQEQKQAQPPVPVPEAKAQVTPPPAPVNNKTNENSKLDYLEKLYE 780
DB 712 AGSALEGDSVQAQAQEQKQAQPPVPVPEAKAQVTPPPAPVNNKTNENSKLDYLEKLYE 771
QY 781 FLNTSYICHKYILVSHSTWNEKILKOYKITKEESKLSLSCDPLDLLFNIONNIPVMYSMF 840
DB 772 FLNTSYICHKYILVSHSTWNEKILKOYKITKEESKLSLSCDPLDLLFNIONNIPVMYSMF 831
QY 841 DSLNLSLSQLFMEIYEKEMVCNLYKLKDNKIKNLLLEAKKVSTSVKTLSSSSMQPLSLT 900
DB 832 DSLNLSLSQLFMEIYEKEMVCNLYKLKDNKIKNLLLEAKKVSTSVKTLSSSSMQPLSLT 891
QY 901 PQDKPEVSANDDTSHSTNLNLSKLNFENILSGKNKIYQELIGQKSSSENFYEKILKDS 960
DB 892 PQDKPEVSANDDTSHSTNLNLSKLNFENILSGKNKIYQELIGQKSSSENFYEKILKDS 951
QY 961 TFYNESFTNFVSKADDDINSNDSEKSKLEEDINKLKTQLSFDLYNKYKLEKLERLFD 1020
DB 952 TFYNESFTNFVSKADDDINSNDSEKSKLEEDINKLKTQLSFDLYNKYKLEKLERLFD 1011
QY 1021 KKTGVCKYKMQIKKTLTLEQESKLSLNNPNKHVLFQNSVFENKKKEABIAETENTLEN 1080
DB 1012 KKTGVCKYKMQIKKTLTLEQESKLSLNNPNKHVLFQNSVFENKKKEABIAETENTLEN 1071
QY 1081 TKILLKHVGLVKNYNGESSPLKTLSEESTQTEDNTASLENFKVLSKLEKLDKNLNEK 1140
DB 1072 TKILLKHVGLVKNYNGESSPLKTLSEESTQTEDNTASLENFKVLSKLEKLDKNLNEK 1131
QY 1141 KKLSSLSGGLHLLIAELKEVKNKNTGNSPSNNNTDNNALSYKFKFPEGTDVATVVS 1200
DB 1132 KKLSSLSGGLHLLIAELKEVKNKNTGNSPSNNNTDNNALSYKFKFPEGTDVATVVS 1191
QY 1201 EGSQDTEQSQPKKPASTHVGAESENTITTSQNVDDDEVDDVIIIPFGESEEDYDGLQVV 1260
DB 1192 EGSQDTEQSQPKKPASTHVGAESENTITTSQNVDDDEVDDVIIIPFGESEEDYDGLQVV 1251
QY 1261 TGEAVTPSIDNTLSKIENEYEVLYLKLPLAGVYRSLLKQLENNVMTFNVNVDILNSRPN 1320
DB 1252 TGEAVTPSIDNTLSKIENEYEVLYLKLPLAGVYRSLLKQLENNVMTFNVNVDILNSRPN 1311
QY 1321 KRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKROKFLSSYNYIKDSIDTDINFA 1380
DB 1312 KRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKROKFLSSYNYIKDSIDTDINFA 1371
QY 1381 DVLGYKILSEKYSKSDLSIKKYINDKQENKYLPLFNNIETLYKTVDNDKIDLFVIHLE 1440
DB 1372 DVLGYKILSEKYSKSDLSIKKYINDKQENKYLPLFNNIETLYKTVDNDKIDLFVIHLE 1431
QY 1441 AKVLNVTYKSNVVEVKIKELNYLKTIDKLADFPKKNNFVGADLSTDYNNHLLTKFLS 1500
DB 1432 AKVLNVTYKSNVVEVKIKELNYLKTIDKLADFPKKNNFVGADLSTDYNNHLLTKFLS 1491
QY 1501 TGMVFENLAKTVLSNLDGNLQGLMNIHQCVKVKQCPQNSGCFRHLDERECKCLLNYK 1560
DB 1492 TGMVFENLAKTVLSNLDGNLQGLMNIHQCVKVKQCPQNSGCFRHLDERECKCLLNYK 1551
QY 1561 QEGDKCVENPNTCTNNNGGCDADAKCTBEDSGSNKKITCECTKPDSPFLPFGIFCSSS 1620
DB 1552 QEGDKCVENPNTCTNNNGGCDADAKCTBEDSGSNKKITCECTKPDSPFLPFGIFCSSS 1611
QY 1621 NFLGISFLLILMLILYSFI 1639
DB 1612 NFLGISFLLILMLILYSFI 1630


```
QY 335 N-----TPN-----TLDDN 344
: : : : :
Db 303 SSASGTSAGAGTVEQANTVASVTVPVSVGQCEASTNPQTAQVQVPVTLTLEEK 362
: : : : :
QY 345 KKEEHEKEKEIAKTAKFNIDSFTDPLBEYLREKKN---NIDISA--KVETKESTE 399
: : : : :
Db 363 KKIAGLQAQKEIAKTAKFNIDLEGIFVDPIELEYFKKEKKESCNLSSTCKKNKASETII 422
: : : : :
QY 400 P--NEYPNGVYPLSYNDI-----NNALNELNSFGDLINFDYTKESKNIYTDNERKKF 452
: : : : :
Db 423 PLTIRYPNGISYPLPENDVNTYKIANNAE---TTYGDTHP-DNTPLTGDATNEQAKDOL 479
: : : : :
QY 453 INEIKIKIEKKIESDKSYEDRSKSLNDITKEVELLNEYDSKFNNDIDLTNPEKM 512
: : : : :
Db 480 IKAIKKIKABEKKLEFLTKYNDKLTNFEQOQTPFKEAAKEFYESFRKNLTSEIFEKF 539
: : : : :
QY 513 MGRKYSYKVEKLTHHNTFASYENSKHLEKLTKALKYMEDYSLRNIVVEKELKYKNLIS 572
: : : : :
Db 540 KTRDRDYMTKK--TELNT--CEYGNTEKELINKLNQNLQYSLRKDIISNEIEVFSNKK 597
: : : : :
QY 573 KIENEIETLVENIKKDEOLFEEKITKDNENKPDKEKILEVSDIVKVOVKVLLMKNKIDELK 632
: : : : :
Db 598 ELQNTINRLAEAVQAKONVLA-----SKOVPLSTLVELQIQKSLLTQKIQEQLN 646
: : : : :
QY 633 KTQILIKNVELKNIHVPNSYKOENKOEYLLIVLKEIDKLVFMPKVESLINEEKNI 692
: : : : :
Db 647 KTEVSLNKAQLKDKLYPKTYGNEGRPEPYLLAVKKEVDRLAQFIPKIESMIKAKERM 706
: : : : :
QY 693 K-----TEGQSDNSPSTEGEITGOATTKPGQOAGSA-----LEG 727
: : : : :
Db 707 EOGPAITGESEVPSPSAESSTDRSQSTSSSSSSSTPAAAESSTATLPEAPAPAEA 766
: : : : :
QY 728 DSVQAOAQEQKQAPVPVPEAKAQVPPAPVANKTENVSKLDYLEKLYEFLNTSYI 787
: : : : :
Db 767 ASPSTEASEETIPTTQETQPSOASSTTPAKPV-----MTKLYLEKLQKLPVFSYS 820
: : : : :
QY 788 CHRYILVSHSTMEKILQYKITEES---KLSSCDPLDLFLNQNIPVYVSNFDSLNN 845
: : : : :
Db 821 CHKVVLLQNSTINKDALKSKYALTSEEDKIRTLKRCSELDVLLAIONNMPTMYSILYESIVD 880
: : : : :
QY 846 SLQOLFMEIYEKWCNLYKLD--NDKIKMLLEA-----879
: : : : :
Db 881 GLQNIYTELEYEKMMHYIKLDENPISIKSLLYKAGVIEPEVPAAPTPTVTPAATQEQOQQ 940
: : : : :
QY 880 --RKVTSVTKLSSSMOPLSLTPQDKPEYSANDTSHS-----916
: : : : :
Db 941 ATPDVQSDAPADSVQSQPETPTVTSITPTVTTEASSSAPGEGTSGEAGASGTGATA 1000
: : : : :
QY 917 -----TNLNNLSKLFEINLS-----LGKKNKIYQELI 943
: : : : :
Db 1001_SNAATPAGTSAGSAGSASNASTTSVDTPPAAAAAASPSTSTPAPAPPAANSQSCNPDGIR 1060
: : : : :
QY 944 GQKSSE-----NFYEKILKSDTPTYNESFNFVKSKADDINSLNDESKRKLEE 992
: : : : :
Db 1061 SRAESEDMPADDFELDNLYKSYLQIQDG--NNTFEFINFKSKELIKALTPPE-KVNQLYL 1118
: : : : :
QY 993 DINKLKTLLSPLDLYNKYKLERLFDKKTGKYKMOIKLTLKEQLESKLNSLNNP 1052
: : : : :
Db 1119 EIAHLKELSEHYDRTSYTKLKERLNKHEQIQLNQRIQRLDLSILKARLLKRRQTLNGV 1178
: : : : :
QY 1053 KHYLQNFSPFNKKKAETAETENTLILKLYKGLVYNGESSPLKLTSEESIQT 1112
: : : : :
Db 1179 FYILNGVYVFNKRRAEQYVDNALKNTDMLKYKARTKYTSEAVPLKTLKSLDLR 1238
: : : : :
QY 1113 EDNYASLENFKVLSKLEGLKONLNKLEKKLSYLSGLHLLIAELKEVKNKNYTNSPS 1172
: : : : :
Db 1239 ESNLYKTEKFRAYSRLERLRLKKNLKNKGERISYVSGLHVFEFEKELIKDKDYTKKNP 1298
: : : : :
QY 1173 ENNTVDNNALESYKFLPEGTQVATVVSSEGSTLEQSQPKPASTHVGAESNTI-----1227
: : : : :
Db 1299 DNAPEVTNAEQYKELLPKGVTVST--PAVAVTTTLAADAAPTPEGAVPGAVPGAVNP 1357
: : : : :
```

```
QY 1228 -----TTSQNVDDDEVDDVILVPIFGESEEDYDLDGQVVTGEAVTPSPVIDNLSK 1376
: : : : :
Db 1358 GAVPGAVPGSGTDRVAGSSVDD-----NED--DDIYQIASQSEDAPEKD--TLSE 1405
: : : : :
QY 1277 IENEYEVLYLKLPLAGVYRSLSKQLENNVMTFNVNVDILNSRNFKNRKNFKNVLESOLIPY 1336
: : : : :
Db 1406 FTRESLYVYTKRLGSGYKSLKHMLREFESTIKEDMTNGLNKNKSQKRNDFLVLSHELDLF 1465
: : : : :
QY 1337 KDLTSSNYVVDKPYKFLNKEKRDKEFLSSYNYIKDSITDINFANDVLGYVKILSEKYSKD 1396
: : : : :
Db 1466 KDLSTKNKYIRNPYQLLNDKDKQIVNLKAYATKGINEDIETTTDGIKFNKKNVLYNTQ 1525
: : : : :
QY 1397 LDSIKYI-----ND--KOGENEKYLPLFLNNIETLYKTVNDKIDLFVHLEAKVLNITY 1448
: : : : :
Db 1526 LAAVEQIOATIEAETNDNTNKEEKKYIPILEDKLGLEYETVIGQABEYSEELQNRLDNYKN 1585
: : : : :
QY 1449 EKSNEVEVKIKELYLTKIQDKLADF-----KKNNFVGIADLSDTDYNNHLLTFLSTGMV 1504
: : : : :
Db 1586 EKAEEFILTKNLEKYIQIDEKLDDEFVEHAENKHIASIA-----LNNLNKSGLV 1634
: : : : :
QY 1505 FENLAKTVLSNLDGNLQGM--LNIS--OHOCV--KOCQPNSSGCFRHLDEREECKLLNYK 1560
: : : : :
Db 1635 GEGESKILAKML--NMDGMDLLGVDPKHVCVDTROIIPKNAGCFRDNNGTEWRCLLYGK 1692
: : : : :
QY 1561 Q--EGDKRCVENPNTNENNGCGDADAKCTEEDSGSKKITCECTKPDSPYPLDFGIFCSS 1619
: : : : :
Db 1693 KGESENTVENNPTCDINNGCGDPTASCQNAESTENSKKICTCKEPTPNAYYEGVFCSS 1752
: : : : :
QY 1620 SNFLGTSFLLMLIYS 1637
: : : : :
Db 1753 SSPMGLSILLIITLIVFN 1770
: : : : :
RESULT 9
MSPL_PLAFD STANDARD; PRT; 233 AA.
AC P13827;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMSA) (Glycoprotein 185) (GP185) (Fragment).
GN MSP-1.
OS Plasmodium falciparum (isolate CDC / Honduras).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87106855; PubMed=3542719;
RA Howard R.F., Ardeshir F., Reese R.T.;
RT "Conservation and antigenicity of N-terminal sequences of GP185 from
different Plasmodium falciparum isolates.";
RL Gene 46:197-205(1986).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
-----
CC EMBL; M14632; AAB59227.1;
DR PIR; A25814; A25814.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 >233 MEROZOITE SURFACE PROTEIN 1.
```

```

FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 233 233
SQ SEQUENCE 233 AA; 26050 MW; 1DE4D31ACAE965 CRC64;

Query Match 14.18; Score 1187; DB 1; Length 233;
Best Local Similarity 95.18; Pred. No. 7.4e-33;
Matches 232; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

Qy 1 MKIIFFLCSEFFFIINTQCVTHESYQELVKKLEALEDAVLFGYSLFOKEKMWLNEGTSGT 60
Dy 1 MKIIFFLCSEFFFIINTQCVTHESYQELVKKLEALEDAVLFGYSLFOKEKMWLNEGTSGT 60
Qy 61 AVTTSTPGSKSVASGGSGSVASGGSGSVASGGSGSVASGGSGNSRRTPNSDSSDS 120
Dy 61 AVTTSTPGSKSVASGGSGSVASGGSGSVASGGSGSGNSRRTPNSDSSDS 108
Qy 121 DAKSYADLKHVRNLYLTIKELKYPQLFDLTNHNMLTLCNDIHGFKYLDGYEETNELLYK 180
Dy 121 DAKSYADLKHVRNLYLTIKELKYPQLFDLTNHNMLTLCNDIHGFKYLDGYEETNELLYK 168
Qy 181 LNFYFDLLRAKLVNDVCANDYCOIPFNKIRANELDVLKLVFGYRKPLDNTKDNVGMKED 240
Dy 169 LNFYFDLLRAKLVNDVCANDYCOIPFNKIRANELDVLKLVFGYRKPLDNTKDNVGMKED 228
Qy 241 YIKK 244
Dy 229 YIKK 232

RESULT 10
MSPI_PLAFN STANDARD; PRT; 281 AA.
ID MSPI_PLAFN
AC P13820;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWSA) (Fragment).
GN MSP-1.
OS Plasmodium falciparum (isolate NF7 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RA "Variation in the precursor to the major merozoite surface antigens
RA of Plasmodium falciparum";
RL Mol. Biochem. Parasitol. 27:291-302(1988).
CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -|- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF
CC 83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19144; AAA29654.1;
CC PIR: B54498; B54498.
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 >281 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 281 281

```

```

SQ SEQUENCE 281 AA; 31226 MW; 6150B76E0F21B299 CRC64;

Query Match 13.08; Score 1096.5; DB 1; Length 281;
Best Local Similarity 77.7%; Pred. No. 8e-30;
Matches 227; Conservative 8; Mismatches 32; Indels 25; Gaps 5;

Qy 1 MKIIFFLCSEFFFIINTQCVTHESYQELVKKLEALEDAVLFGYSLFOKEKMWLNEGTSGT 60
Dy 1 MKIIFFLCSEFFFIINTQCVTHESYQELVKKLEALEDAVLFGYSLFOKEKMWLN-----E 56
Qy 61 AVTTSTPGSKSVASGGSGSVASGGSGSVASGGSGSVASGGSGNSRRTPNSDSSDS 112
Dy 57 EITT-----KG--ASQSGASQSGASQSGASQSGSGTSGPSPSSRNTLPSRN 109
Qy 113 -----PSDSSSDSAKSYADLKHVRNLYLTIKELKYPQLFDLTNHNMLTLCNDIHGFKY 166
Dy 110 TSSGASPPADASDSDAKSYADLKHVRNLYLTIKELKYPQLFDLTNHNMLTLCNDIHGFKY 169
Qy 167 LIDGYEETNELLYKLVNFEYDILLRAKLVNDVCANDYCOIPFNKIRANELDVLKLVFGYRK 226
Dy 170 LIDGYEETNELLYKLVNFEYDILLRAKLVNDVCANDYCOIPFNKIRANELDVLKLVFGYRK 229
Qy 227 PLDNIKDNVGMKEDYIKKNKKTITENINELIEBSKKTIDKNKNATKEEKKKL 278
Dy 230 PLDNIKDNVGMKEDYIKKNKKTITENINELIEBSKKTIDQKNADNEEGKKKI 281

RESULT 11
YD86_SCHPO STANDARD; PRT; 1957 AA.
ID YD86_SCHPO
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C1F3.06c in chromosome I.
GN SPAC1F3.06c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkert G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle V.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; 270690; CAA94624.1; -
CC KW Hypothetical protein.
CC SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 6.0%; Score 501.5; DB 1; Length 1957;
Best Local Similarity 22.2%; Pred. No. 1.5e-09;
Matches 352; Conservative 253; Mismatches 617; Indels 361; Gaps 67;

QY 105 SGNRRTPNSDSDAKSYADLKHVRNYLLTITELKYPQLFDLTNNHMLTCDNIHGF 164
DB 105 SGNRRTPNSDSDAKSYADLKHVRNYLLTITELKYPQLFDLTNNHMLTCDNIHGF 164
DB 162 SNNKDKNTSVTLTSEEDVYFQKLTN-MESNFAKQSEAYDLSRQLTATEKDKK 220
QY 165 KYLDGVEETNELLYKLNFYDLRLAKLNDVCANDYQIPFNLIKIRANELDVLKLVFGY 224
DB 165 KYLDGVEETNELLYKLNFYDLRLAKLNDVCANDYQIPFNLIKIRANELDVLKLVFGY 224
DB 221 E--KDYETIKEDV-----SSIKASLAEEQASN-----KSLRGEQERLEKLLVSS 262
QY 225 RKLPLDNKQVGMEDYIKKNTKIENINELIEESKTTIDKKNATKEBEKKLYQAQYD 284
DB 225 RKLPLDNKQVGMEDYIKKNTKIENINELIEESKTTIDKKNATKEBEKKLYQAQYD 284
DB 263 NKTVSTLR-----QTSNLAECRTQ---EKLE-----KCAINEEDSKLLEELKHN 306
QY 285 LSYNKOEEAHLNLSVLEKRI-----DTLK-KNENIKELLDKINEIKNPPANS 333
DB 285 LSYNKOEEAHLNLSVLEKRI-----DTLK-KNENIKELLDKINEIKNPPANS 333
DB 307 VANYDAIVHKKLIEDLSTRISFDFNLSKERTSLIKNEKLEKLL----- 352
QY 334 GNTPTNLLD---KNKITEEKEIKEIKTI-----KENIDSLFTDPLEYIYREKN 383
DB 334 GNTPTNLLD---KNKITEEKEIKEIKTI-----KENIDSLFTDPLEYIYREKN 383
DB 353 RNTIGSLKDSRTSNSQLEEREWELVESNRTIHSQLTDAESKLSSEFQENKSLGIDEYQ 412
QY 384 KNTDISAKVETKESTNEPVGVTPLSYNDINNALNELNSFGDL-----INFPDITKEP 439
DB 384 KNTDISAKVETKESTNEPVGVTPLSYNDINNALNELNSFGDL-----INFPDITKEP 439
DB 413 NNLSSKDKMYQVSSQLEEARSLAH-----ATGKLAIEINSERDFQNKIKDFEIQD 466
QY 440 SKNIYDNERKKFTNEIKETKEKK-----IESDKSYEDRSKSLNDITKEY- 488
DB 440 SKNIYDNERKKFTNEIKETKEKK-----IESDKSYEDRSKSLNDITKEY- 488
DB 467 LRACLNSS-----NELKESALIDKQDLNLRQIKQKVSSTQSSLSQSLQDIL 521
QY 489 -EKLLENIYSKNN-----NIDLTFEKMGRYSYKVEK-----LTHNTFASYENSKHN 539
DB 489 -EKLLENIYSKNN-----NIDLTFEKMGRYSYKVEK-----LTHNTFASYENSKHN 539
DB 522 NKKKKHEVYESQLNELKGELOTEISNSELHLSQSLTLAAEKAATAVATNNELSEKNSLOT 581
QY 540 L-----EKLTKALKYMEDYSLRNVVEKELKYKLNLSKLENELETIVENIKKDEQL-- 592
DB 540 L-----EKLTKALKYMEDYSLRNVVEKELKYKLNLSKLENELETIVENIKKDEQL-- 592
DB 582 LCNAFQELAKSVQKLENQNFSSLDTSFKLNEHSHOELNNHOTITKOLKDTSSKLOQ 641
QY 593 -----FEKK-----ITKDNKPKDEKILEVSDIVKQVQVLLMKNKIDELKKTOLILKNVE 642
DB 593 -----FEKK-----ITKDNKPKDEKILEVSDIVKQVQVLLMKNKIDELKKTOLILKNVE 642
DB 642 QLQERANFEQKESTLSDENNDRTLKLLKLESNKSLLIK--QEDVDSLEKNITQTLKE-D 697
QY 643 LKINIHPNSYKQENKQEPYIIVLKEIKDKLVFMPKVESLINEEKNNIKTEGQSD--N 700
DB 643 LKINIHPNSYKQENKQEPYIIVLKEIKDKLVFMPKVESLINEEKNNIKTEGQSD--N 700
DB 698 LRKSEALREFSKLEAKN-----LREVIDNLK--GKHETL--EAOENLHLSLSDAKN 745
QY 701 SEPSTEGEITGOATTPGQAGSALGSDSVQAQAEQKQAPPPVPPVPAKAQVTPPPA 760
DB 701 SEPSTEGEITGOATTPGQAGSALGSDSVQAQAEQKQAPPPVPPVPAKAQVTPPPA 760
DB 746 TNAILSELF-----KSESDVKRLTANVETLTQDSKAM----- 778
QY 761 PVNKNTEVSKLDYLEKLYEFLNTSYCHKYILVSHSTNNNEKILQYKVTKEESKL-SS 819
DB 761 PVNKNTEVSKLDYLEKLYEFLNTSYCHKYILVSHSTNNNEKILQYKVTKEESKL-SS 819
DB 779 -KOSFTSLVNSYOSINLYHELDRDH-----VNNQSQNTLL-----ESESKLTD 823
QY 820 CDPL-----DLLFNION-----NIPWYSMFDSLNLSLSQPMVEYKEMVCNLYKLKD 868
DB 820 CDPL-----DLLFNION-----NIPWYSMFDSLNLSLSQPMVEYKEMVCNLYKLKD 868
DB 824 CENLTQONTMTLIDNVQKLMKHVHNSKVSSELKEVNGKLS---LDLKNLRSSLN--AISD 879
QY 869 NDKIKNLEAKKVSIVKTLSSSSWQPLSLTPQDREYSANDTSH--STNLLNSLKLF 926
DB 869 NDKIKNLEAKKVSIVKTLSSSSWQPLSLTPQDREYSANDTSH--STNLLNSLKLF 926

DB 880 NQILTLQALBSKNYDSLEQESQALNSGLSKLSBAEQKOLLHTENEELHRLDKITGLKIE 939
QY 927 ENILS--LGNKNKIYQELIGOKSSSENFYEKILKDSDFYNESEFTNFVKSADDINSLNDES 985
DB 940 ESKSDGLGKLTARQEISEINLKEENMSQ-----SOAITTS-VKSKLDE-----TLS 983
QY 986 KRKKLEEDINKKTKTLQLSFDLYNKYKLLERLFLDKKKTGVK-----YKMQIKKLTLLKEQ 1041
DB 984 KSSKLEADIEHLKKNVSEVEVERNALLASNERLMDLKNNGENIASLOTIEKKRAENDD 1043
QY 1042 LESKLSLNNPKHVLFQNFVFFNKKKEAEIAETENTLENTKILLKHKGLVLYKYNBESSP 1101
DB 1044 LQSKLSVSVSEYENLLLISSQTNKLSLEDTNQLKYTEKNVQKLLDEKQDNVELEELTSK 1103
QY 1102 LKTLSPESIQTDENYASL--ENFKVLSKLEGLKLDNLEKKLSYLSLSSGLHLLIAELKEV 1160
DB 1104 YGLGGENAQIKDELLALRRKSKQHDLCANFVDDDLKEKSDALEQLTNEKNELIVSLEQ- 1162
QY 1161 IKKNNTTGSPPSENNTDVNNNALESYKFKFLPEGTDTAVTVVSESGSDTLEQSQPKKPASTHV 1220
DB 1163 ---SNSNEALVEERSDLANRLSDMKKSLSDSDNVISVIR---SDLV----- 1203
QY 1221 GAESNTITTSQNVDDDDV-----IIVPIFGESEEDYDDLGQVVTG----- 1262
DB 1204 -----RYNDELTLKKDKSLSTQYSEVQCDRDDLLDLSLKGCEESFNKYAVSLR 1252
QY 1263 EAVTPSVIDNLSKIENEYEVLYKPLAGVYRSLKQLENNVMTFN-VNVKDI-LNSRFN 1320
DB 1253 ELCTKSEIDVPVSEILDDNFVFNAGNFSELSRLTVLSLENYLDAFNQVNFKKMELDNRLT 1312
QY 1321 KRE-NFKNVL-----ESD--LIPYKDLTSSNYVVKDYPK-FLNKE----- 1356
DB 1313 TTDAEFTKVVADLEKLQHEHDDWLIQRGDLEKA---LKDSEKNFLRKEAEMTENIHSLEE 1369
QY 1357 -----KROKFLSS-----YNYIKDSID---TDINFANDVLGYKYL----- 1389
DB 1370 GKEETKKEIAELSRLEDNQLATNKLKNQDLHLNKEDEVLKKEKESLIISLEESLSN 1429
QY 1390 -SEKYKSDLDS---IKKYINDKQGENEKYLPFLNNIETLYKTVDKI-----DL 1434
DB 1430 QKQKESLLDAKNEHMLDDTSRKNS---LMKEITESNSSLDDKSFELASAVEKLGAL 1486
QY 1435 FVHLEA-----KVLNITYEKSNV-EVKIKELNLYKTIOQKLDADFKANNFVGIADL 1485
DB 1487 QKIHSELSLIMENIKSQLQEAKEIKQVDESTIOELHEITASKNNYEGKLNKDXDIIRD 1546
QY 1486 STDYNH-NLL-----TKFLST 1501
DB 1547 SENIEQLNLLAEKSAVKRLST 1569

RESULT 12

USOL YEAST	STANDARD;	PRT;	1790 AA.
ID	USOL YEAST		
AC	P25386;		
DT	01-MAY-1992 (Rel. 22, Created)		
DT	01-MAY-1992 (Rel. 22, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Intracellular protein transport protein USOL.		
GN	USOL OR INT1 OR YDL058W.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=X2180-1A;		
RX	MEDLINE=91185402; PubMed=2010462;		
RA	Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,		
RA	Yamasaki M.;		
RT	"A cytoskeleton-related gene, usol, is required for intracellular		
RT	protein transport in Saccharomycetes cerevisiae."		
RL	J. Cell Biol. 113:245-260(1991).		


```
Db 1005 QNNKEVEYKSEYKSLQNDLQDTIYANTAOQNNYQEQELQKHADVSKTISELREQ 1064
QY 855 YEMVNCVLYKLDNDKIKKLEAKKVSVK-----TLSSSMOPLS-----LTPQ 902
Db 1065 HTYKGQVKTUWL-SRQOENALNENKSWSSQESLLEQLDLSNRTEDUSSQNKLLYDQ 1123
QY 903 DKPEVSANDTSHSTN---LNNSLKLPLNTSLGKNKNIY--QELIGOKSSENFYEKI-- 955
Db 1124 IQIYTAADKEVNSTNGPGLNNIL-----ITLRREDIILDTKVYVAERDAKMLRQKISL 1177
QY 956 ----LKSDTFYNESFT-----NFVKSADDINSL-----NDSKR 987
Db 1178 MDVELODARTKLDNSRVEKENHSHSIIOQHDDIMEKLNQMLNLLRESNITTLRELENNNNKK 1237
QY 988 KLEEDINKLKKTL-----OLSPFLYLNK-YKLKLERLFDKKTKVGYKMOIKKLTLL 1038
Db 1238 KEIQSELDKLQNVAPIESLTAIKYSMOEKOELKL-----AKEVHRWKRSODILEK 1292
QY 1039 KEQLES-----KINS-LNNPKHVILQN-----FSVFNK-KKEAE-----IAE 1073
Db 1293 HEQLSSSDYKLESEITENLKEELENKROGAEEKFNRLRQAOERLKTSLQSDSLTE 1352
QY 1074 TENTLENTKILLKHGYKLVKYNGESSPLKTLSEESIOTDDNYASLENFVYL---SKLEG 1130
Db 1353 QVNSLRDAKNVLEN-----SLSEANARIEE---LQNAKVAQGNQOLEA 1392
QY 1131 --KLKDN-----LNLEKKLSYLS--SGLHHLIAELKEVIK-----NKNYTGNSPS 1172
Db 1393 IRKLQDAEAKSRELQAKLEESTSTVESTINGLNEELTLKEIEERQORQOQLOQATSAN 1452
QY 1173 ENNTDVNNALESYKKFELPCTDVATVYVSESGSDTLE---QSQPKPASTHVGAESNTITT 1229
Db 1453 EQN-DLSNIVESMKKSFE--DKIKFIKEKTQEVNEKILEAQERLNQPSNINNEEIKKKW 1509
QY 1230 SQNVDDVEDVDVIVPTFGESEEDYDGLGVVTVGEAVTPSDINILSKIENEYEVLYKPL 1289
Db 1510 ESEHEQVQSOKI-----REAEALAKRIRLPTTEKIN-KIIEKKKELEKEFEKEVEERI 1563
QY 1290 AGVRS-----LKKOLENNVMTFNVVKDILNSRKRENFKNVLESIDLIPYKDLTSSN 1343
Db 1564 KSEQSGEIDVLRKQLEAKVQ---EKQKELENEVKK-----LQEEI---KDVPHSS 1610
QY 1344 YVVKDPYKFLUNKRDRKFLSSYNIKDSIDTINFANDVLGYKILSEKYSKDSLSIKKY 1403
Db 1611 HISDD-----ERDK-----LRAEIESR-----LREEFNELQAIKKK 1642
QY 1404 INDKQGENEYKLPFLNNIETLYKTVDKIDLFIHLEAKVLTNYEKSNEVEVAKELNLYL 1463
Db 1643 SFD-EGKQ-----QAMMKT-----LLERK-----L 1662
QY 1464 KTODKLADFKKNNFVGJADLSTDYNHNNLLTKELSTGWFENLAKTVLSNLLDGNLQ 1523
Db 1663 AKNESQLSETKQSAE-----SPPKSVNVQNPGLGLPRKIEENSFPNPLSG----- 1711
QY 1524 MLNISQHCYKQCPQNSGCFRHLDEREECKCLLNTKQEGDKCVENPNP 1572
Db 1712 -----EKLKLNKSSSGGFNFTSPSPNKHQNDNDKRESLANKTDP 1754

RESULT 15
RBP2_PLAVB
ID RBP2_PLAVB STANDARD; PRT; 1251 AA.
AC Q00799;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 2 (Fragment).
GN RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RA MEDLINE=92315338; PubMed=1617731;
RX Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites.";
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M88098; AAA29744.1; -.
KW Malaria; Receptor; Membrane.
FT NON_TER 1
FT NON_TER 1251
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 5.4%; Score 457.5; DB 1; Length 1251;
Best Local Similarity 21.1%; Pred. No. 2.5e-08;
Matches 306; Conservative 228; Mismatches 458; Indels 455; Gaps 71;

QY 272 EEEKKLYQAQYDLSYNNQLEEAHNLISVLEKRIIDLTKKNENIKELLDKINIKPPPA 331
Db 1 EFDKERVKDTSPDEK---KQSIKAY-----EKMGTLEK-----LEKMDDEKN--- 41
QY 332 NSGNTNTLLDNKKKEEHEKEIKIAKTIKFNID---SLFTDPLEYLYLRK-----NK 384
Db 42 -----IKVEEAQIQYKRIF-----IDHDVNLMDNEVEKSKVIMEKIELYKK 84
QY 385 NIDISAKVETKESTEPNEYPNGVT-----YPLSYNDINNALNELNSGDLINPDYTKPS 440
Db 85 EID-----EIKQKT--NEYKQDTSNFYYTEQYNSATQSKAKIEQFINIATTKGSDTS 137
QY 441 KNIYTDNERKKFFINEKEKIEKKIESDKKSYEDRSK---SLNDITREYKLLN--EI 495
Db 138 QDI---NE-----LESKEEVHKNLQLVKQESNMEERKQILSMKDLL-----ILNSET 185
QY 496 YDSKFNNTDLTNFERNMGKRYKVEKLTHTHTFASYENSKHNLEKLTALKAYMEDYSL 555
Db 186 IAKEISNT-----QNALGFRENAKTKLAKTDELQ----- 216
QY 556 RNTVVEKELKYKNLISKIENIEITLVENIKKDEBOLFEKKITKDKENKPKDEKILEVSDIV 615
Db 217 RVAAMTEEAKAHK-----NNIDIALEDAQID-----TEVSKIE 249
QY 616 KVQVQVLLMNKIDELKTKTLILKNVELKHNHVPNSYKOENKQEPYLLIVLKEIDK 675
Db 250 QIMRE---TMNKDEIKSYLSEIKYKDKCTTEISNKRKGDK-----IEFLE 294
QY 676 VPMKPVES-----LINEKKNIKTEGQ-----SDNSEPTEGEITGOATTKPGQOAGSA 724
Db 295 KFKPNEESNKNVINEINENIRNSEQYIKDIADAQKASTKVELFKHET---TISNIF 351
QY 725 LEGDSVOAQAEQKQAPPPVPVPEAKAQVPTPPAPVNNKTNVSKLDYLEKLYEFLNT 784
Db 352 KESEILGVETKSQKI-----NKAEDIMK--EIER----- 379
QY 785 SYICHYILVSHSTMNEKILKQYKIKTEESKLSKSCDPLDLFNIGNNIPVMY-SMFD 843
Db 380 -----HNSEITQTVKGFQENLKL-----NEPHNYDNAEDEL 411
QY 844 NNSLSQLEWIEYKEMVCNLYKLKDN-DKIKNLEBAKKVSTSVKTLSSSSMOPLSLTPQ 902
Db 412 NNDKSTNAKVLIE---TNLESVKHNLSEITNKGGEKIYSKAKDI---MQIKATSE 463
QY 903 DKPEVS---ANDTSHSTNLNLSKLFIENILSLGKNK-----NIYQELIGQKSEN 950
```

```
Db 464 NTAETLEKVKDDQSNVYVNLNGIUTERNLIVTEKRNLRGIDSTITNIEGAL---KESKG 520
QY 951 FYE-----KILKDSDTFYNESTNEVSKADINSIND-ESKR 987
Db 521 NYEIGFLEBEIGKRNKRLKVDIRKTSINSTVGNFSLNFDFLNOYDFNKINDYENKM 580
QY 988 ----KLEEDINKLKTQLSFDL---YNNKYK-LKLE-----RLFDRKKKTGVKYMQLK 1033
Db 581 GEIYNEFEGLSKISENLNASENTSIDYNSAKTLRLLEAGKEKVNLLNKEEANKYLRDVK 640
QY 1034 KUTL-----LKEQLSKNSL-----NNPKHVLFQNSVFFNKKKEAEIAE-- 1073
Db 641 KVESFRIFNMKRESLD-KINEMIKKQLAVNEGHGNVQLVENIKELVDENNLSDILKA 699
QY 1074 -----TENTLEN-TKILLKHVGLVYNGESSPLKTLRSIESIQTDNYASLENF 1122
Db 700 TKNEBIEQIHTLTKANKAKTILGHVDTSKIVGKITPELALTE----- 744
QY 1123 KVLKLEGLKLDNLEKKKLSYLSGLHHLIAELKEVINKNNTGNSPSENNTDYN-NA 1181
Db 745 -LLGDAKLAQELKPEK-----NNVLETEMNSKN-----TNELDVHKNI 785
QY 1182 LSEYK---KFLPEGTDAVTVSESG-----SDTLEQSPKPA-ST 1218
Db 786 QDAYKVALEILAHSDIEDTKORDSKSLIEMGNQIYLVVLIINQYKNKISSIKSKEEAVSV 845
QY 1219 HWG-----AESNTITTSQNVDEVDVLIIVPIFGESEEDYDGLGVVTGEAVTPSVIDN 1272
Db 846 KIGNVSKKHSLSKITS---DKSYDNIITAI-----EKQTELONLRSFTOEK-TNINSDS 897
QY 1273 ILSKIENEYEVL--YLKPLAGVYRSLSKQLEN--NYMTFNVNVVKDILNFRNFKRENFRNV 1328
Db 898 KLEIKRTFESLKNALKTLEGEVNALKASSDNHEHVQSKSEPVNPAL-SEIEKEET---- 952
QY 1329 LESDLIPYKDLTSSNVVVDKPKFLNKRDRKFLSSYNIKDSI-----DPTDI--NFPAN 1380
Db 953 -----DIDSLNATLDE---LLKGGRTCEVSRKLIKDTVTKEISDDTELINTIBK 999
QY 1381 DVLGYKILSEYK---SILDSIKKYINDKQGENEKYLPF-----LNNIETLYKTV 1428
Db 1000 NVKAYLAYTKKYEDTVQDVLNLFNEHFNFKQVSNHEPTNFDKSNKSSELTAKVDSKTI 1059
QY 1429 NDKIDLVFHLPAKVLNITYEKSNEVKIKEL-----NYLKTIOQDLADFKNNNFVGIAD 1484
Db 1060 ISLKGVIIE-----VNENTEMNTIESSAKEIEALYNELNKKKTSLSNEIYQTSNEVKLQE 1114
QY 1485 LSTDYNNHLLTKFLSTGVGFENAKTVLSNL-----LDGNLQGLMNLISQH- 1530
Db 1115 MKSNAD-----KYIDVSKIFNTVLDPTQKSNIVTNOHSINNRYKDKLKGKLQELIDADSSF 1168
QY 1531 --QCVKK 1535
Db 1169 TLESIKK 1175

RESULT 16
GENE HUMAN
ID CENE HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
```

```
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RT Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RT Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 215005; CAA78727.1; -.
CC PIR; S28261; S28261.
CC DR HSSP; P17119; 3KAR.
CC DR Genew; HGNC:1856; CENPE.
CC MIM; 117143; -.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
CC Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
CC Cell cycle; Centromere.
CC DOMAIN 1 335 . KINESIN-MOTOR.
CC DOMAIN 336 2471. COILED COIL (POTENTIAL).
CC DOMAIN 2472 2663. GLOBULAR (POTENTIAL).
CC NP_BIND 86 93. ATP (BY SIMILARITY).
CC SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;
CC -----
CC Query Match 5.3%; Score 443.5; DB 1; Length 2663;
CC Best Local Similarity 18.4%; Pred. No. 1.6e-07;
CC Matches 345; Conservative 349; Mismatches 617; Indels 559; Gaps 82;
QY 126 ADLKHVRNVLTLIKELK---YFQLFDLTNHLT-----LCDNIHGRKY 166
Db 415 AKRRKRVTCWLGKINKMKNSVDQFNPTNITTTTHKLSINLLRDEISVCSESDFSN 474
QY 167 LIDGVEEI-----NELLYKLNFFYDL--LRALNDVCANDYQCI-----PFNLKIR-ANE 213
Db 475 TIDTSEIETWNPATKLLNQENIESELNSLRADYDNLVL-DYQLRTEKEEMELKLEKND 533
QY 214 LDV-----LKKLVFG---YRKPLDN-----IKDNVKG 237
Db 534 LDFEALERKTKKQDQMQLIHEISNLKLVKHREYVNDQLENELSSKVELLREKEQIKK 593
QY 238 MEDYTKKKKKTIENTIN-----ELIEESKK-----TID-KKNKATKEEKKKL 278
Db 594 LQEYIDSQR--LENIKMDLSYLSIESIEDPKQMKQTLFDFAETVALDAKRESAFRSLENLE 651
```


RL J. Bacteriol. 175:7918-7930(1993).

CC - FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH

CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS

CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW

CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS

CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY

CC SIMILARITY).

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U39701; AAC71437.1; --

DR EMBL; U02165; AAD12447.1; --

DR TIGR; MG218; --

KW Cytoadherence; Structural protein; Coiled coil; Complete proteome.

FT DOMAIN 28 838

FT COILED COIL (POTENTIAL).

FT DOMAIN 914 1591

FT COILED COIL (POTENTIAL).

FT DOMAIN 1632 1723

FT COILED COIL (POTENTIAL).

FT DOMAIN 1777 1804

FT COILED COIL (POTENTIAL).

SQ SEQUENCE 1805 AA; 216252 MW; 110933AF173284FD CRC64;

Query Match 5.2%; Score 440; DB 1; Length 1805;

Best Local Similarity 19.7%; Pred. No. 1.4e-07;

Matches 334; Conservative 301; Mismatches 566; Indels 492; Gaps 79;

QY 133 RNYLTI-----KELYPOLFDTLNMHMLTCDNIH-----GPKYL 167

DB 190 KNYLNVLDQYNELDQLENOKRLLSIEYENTYRELVSADNQLQVYENIDQNOIQFHHQ 249

QY 168 IDGY-EEINELLYKLF-----YFOLLRAKLDVWCANDYCOIPFNKIRANEL-DVLK 218

DB 250 YTYRDELSQLERKQLTQKQELVDKESALRVKIDDA-----DFYNARLAELDDVAK 301

QY 219 KLVGYRKLPLONI-KDNVGMEDYIKKKKTIENELIE-----ESKTTIDKKNKATKE 272

DB 302 QLSF-----QDGITKONAHQVEDKLVALNKEKDRLENTQKEAFFNLQRSALIDNK----- 351

QY 273 EEKKLYAQVDLSYINKQLEBAHNLISVLEKRIDTLKNE-NIKELDKINEIKNPPPA 331

DB 352 -----LQOENELFAKHLHQOEFEE--QKQSDSLKLETEYKALQHKINEPKN----- 397

QY 332 NSGNTPTLLDNKKIEHEKEIKEIAKTIFKNIDSLFTDPLEYLYLRXNKNIDISAK 391

DB 398 ESATKSEELLNQEERLEFKRRE-----IDTLTQ-ASLEYEHORESSQLLKDQ 445

QY 392 VETKESTEPNPGVTPYPLSYNDINNALNELNSFGDLINPDYTKPSKNITYDNERRK- 450

DB 446 NEVKHQNLEY-----AKKELDKERNLL-----DOQKKV 475

QY 451 --KFTNEIKETIKIEKKIES---DKSYEDRSKSLNDITKEYEKLNEIVDSKPNND 505

DB 476 DSEAFOLKEKVAQERKELEELYVKKQDOKE--NELL-PFEKQLKQ-HOAPFENELE 531

QY 506 LTNFEKMGKRYSYKVEKLTHNTFASVENSXHNLEKLTALKYMEDYSLRNIVVEKELK 565

DB 532 AKQELFEAK---HALER-----SPIKLEDEKDLN--TKAOTIANEFS-----QLK 573

QY 566 YKKNLSIKIENIEFTLVENIKKDEPQLEKFKITKDN-----KPEKILEYSD--- 613

DB 574 TDKSADFELMLQNEYENLOEQEKQLFQERTYFPFNAVLNRLQOQREELLOQKETILD 633

QY 614 -----IVKQVQVLLMKNKIDELKKTQLILKNVELKHNHVPNS 653

DB 634 QLTKSFEQERLINQREHKELVASVEKQELKGLKQDFSQTSLSNASKNLAEREMAIKKE 693

QY 654 KOENKQEPYLLVLKKEIDKLVKMPKVESL---INEKKNIKTGG-----SONSEPS 704

DB 694 KEIATEKQLL-----NDVNNNAEVIQADLAQLNQSLSNORSELQNAKQRIADFNHDSLKKL 749

QY 705 TEGEITGOATTKPGQOAGSALGDSVQAQA-----QEQKQA-----QPPVPVPVPPA 751

DB 750 NEVELSLQKRLQELQTLLEANQKHQSYQOAYFEGELDKLNREKQAFNLNRKQTMEVDAL 809

QY 752 KAOVTPPPAPVNNKTENVSKLDYLEKLYEFLNTSYCH-----KYLL 793

DB 810 KORLSD-----KHOALNQQAELDRKTHE-LNNAFLNHADADQKSLQDQATVKTQKLD 863

QY 794 VSHSTMNEK-----ILKQVKITKEESKLSSCD-PLDOLLFNION 831

DB 864 LERSALLEKQREFAENVAGFRHWSNKTSQLOKIYELTKKQSEQOTKTELKIAF---S 920

QY 832 NIPVMYSMPDLSUNSLSQLEMEIYKEMVCNLYKLKDNNDKIRNLLBEAKKSVSTKLSS 891

DB 921 DLQKDYQVFELQKD--QEFQRQIEAKQ-----RELDKL-----AEK-NNQVKLELD 962

QY 892 SSMQPLSLTPQDKPEVSAN-DDTSHSTNL-----NNSL-----KLEFNILSLG---K 934

DB 963 NRFQALQNOQKQDTVOAQLELEREHQHNLQEQAFNOANESLLKQREQLTKKIQAFHYELK 1022

QY 935 NKNIYOELIGOKSSSENFVEKILKQSD-----TFYNESFTNFVKSRADDINSND----- 983

DB 1023 KRNQFALAKGRFLFAKEQDOQRKQDEINWRFPQKPEYTFDFEAKKRELEELKIRSL 1082

QY 984 -----ESKRKKLEED---INKLKTQLSFDLYNKKYLKLERLFDKXKTVQKY---KMOI 1032

DB 1083 QSNVELERKREKLATDFTNLNKVQHNTQINRDQNS---QIROFLLEKRNKQFQSEANA 1139

QY 1033 KKLTLKQLESKLSLNNPKHVLQNFVSFFNNKKEAEIAETENT---LENTKILLKHY- 1088

DB 1140 KKAFLIK-RLRSFASNLKLQKEALAIQLEFQDKRDEQOQKKELOQATLQLEQKFEKQNF 1198

QY 1089 ----KGLV-----KYNGESSPLKTLSEESI-----QTDENVASLE 1120

DB 1199 TEKQRLVAIKQCEKLSDEKALKQKLVELANLSQTYLANKNKAEYSOQOQKQYTNLL 1258

QY 1121 NPK-----VLSKLEGLKDNLNLEKKKLSYLSGLHHLJIAELKEVYKNNK 1164

DB 1259 DLKENLERTKQDLKKHRSIFARLT-KFANDLRFKKQLLKAQRIVDKDNRLKERNEL 1317

QY 1165 NYTGNSPENNTDVNNALESYKFLPEGTDVATVVSSESDTLEQSOPKPKPASTHVGAES 1224

DB 1318 HFLSNETERKRAVLEDQISYFEKQKQATD-AILASH-----KEYKKEGELQKLLVEL 1370

QY 1225 NTITTSQNVDDDEVDDVVIIPFGESEEDYD-----DLGOVVTGEAVTPSVIDNLSKI 1277

DB 1371 ETRKTKLND-----FAKFSROREFEFNORLKLLELQKTLQOTQNSNNFKTKAIQEI 1422

QY 1278 ENEY-----EVLYLKP-----LAGVYRSLKQLENMNTFNVNKD-----ILNSR 1318

DB 1423 ENSYKRGMEELNFQKKEFDKNKSRLYEYFRKMRDEIRKESQVKLVLEKQKORANLLEAQ 1482

QY 1319 FNKRNFNKL---ESDLPIYK-----DLTSSNVVVKDYPKFLNKEK-----RDKPL 1362

DB 1483 ANKLNIERTIDFKEKELKAFKDKVDQDIDTSTNORKELNELNKLQOQLIERERA 1542

QY 1363 SSYNYIKOS-----IDTDINFAND-----VLGYKYLSEKYKSDL-----DSIKK 1402

DB 1543 NS-----KOSLLNKKLTETIKRQLHDKEMRVRLVLDNRKLAEQKYQTEINRLRTQTFDSEKQ 1598

QY 1403 -----YINDKQGEN-----ERYLPP-----LNN 1420

DB 1599 DIKNFPPFLFKINGNDMAFFLYLWLYPOOKODNTLOIRQLFEQOLQPMQOQYENELNE 1658

QY 1421 IETLYKTVNDKIDLPVHLEAKVLNYYTEKSNVYKIKELNYLKTIDQKLADFKNNNPV 1480

DB 1659 LRRQRNLLEKKLD--QIOLESOLNNKQSEFSKVESMMEKL--LEKTESRLNDFDQKINY- 1713

QY 1481 GIADLSTDYNNHN 1493

DB 1714 ----LTKVYNQHN 1722

RESULT 18

GOG4.HUMAN STANDARD: PRT: 2230 AA.
ID GOG4.HUMAN Q14436; Q13270; Q13654;
AC Q13439; Q14436; Q13270; Q13654;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa
golgin) (Golgin-245) (72.1 protein).
GN GOLGA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=96215236; PubMed=8626529;
RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT "Molecular characterization of trans-Golgi p230: a human peripheral
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif.";
RL J. Biol. Chem. 271:8328-8337(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
RT protein containing a granin signature.";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balague C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC -!- FUNCTION: MAY PLAY A ROLE IN VESICULAR TRANSPORT FROM THE TRANS-
CC GOLGI.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
CC GOLGI MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS: At least 4 isoforms: 1 (shown here), 2, 3
CC and 4; are produced by alternative splicing.
CC -!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND
CC IN HEPATITIS B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U41740; AAC50434.1; -
DR EMBL: X82834; CAAS8041.1; -
DR EMBL: U31906; AAC51791.1; -
DR EMBL: X76942; CAAS4261.1; -
DR Genbank: HGNC:4427; GOLGA4.
DR MIM: 602509; -
DR InterPro: IPR000237; GRIP_domain.
DR Pfam: PF01465; GRIP; 1.
KW Golgi stack; Antigen; Coiled coil; Alternative splicing.
FT DOMAIN 133 237 COILED COIL (POTENTIAL).
FT DOMAIN 276 1011 COILED COIL (POTENTIAL).
FT DOMAIN 1033 1214 COILED COIL (POTENTIAL).
FT DOMAIN 1259 2152 COILED COIL (POTENTIAL).
FT VARSPLIC 2154 2185 TPYKGNLYHTDVSIFGPEYELRKVLFEY -> HLTKV
FT AICTIRMSHLENPLNSICEKFLSI (IN ISOFORM
FT 2).

FT VARSPLIC 2186 2230 MISSING (IN ISOFORM 2).
FT VARSPLIC 2103 2109 MISSING (IN ISOFORM 3).
FT FTSPRGIF -> SWLRSS (IN ISOFORM 4).
FT CONFLICT 188 188 R -> K (IN REF. 3).
FT CONFLICT 220 220 Y -> H (IN REF. 3).
FT CONFLICT 276 276 T -> A (IN REF. 3).
FT CONFLICT 584 584 K -> E (IN REF. 3).
FT CONFLICT 628 628 T -> A (IN REF. 3).
FT CONFLICT 630 630 K -> E (IN REF. 3).
FT CONFLICT 682 682 K -> N (IN REF. 3).
SQ SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;
Query Match 5.0%; Score 424; DB 1; Length 2230;
Best Local Similarity 19.4%; Pred. No. 5.8e-07;
Matches 343; Conservative 303; Mismatches 546; Indels 572; Gaps 81;
QY 26 QELVKKLEALE-----DAVLGTGSLFOKEKMWLNESGTAVTSTPGSKGSVAS 75
DB 137 EQLIQLRLRMERSLSYRGKYSELVTAYOMLOREKKL-----QGLS- 179
QY 76 GSGGSGVASGSGVASGSGVSGGSGNSRRTPNSDSSSDSAKYADLKHVRNY 135
DB 180 -----QSQDKSLRRIAELELQMD 199
QY 136 LTIKELKYPQLFDLTNHLMTLCDNIHGFYLDIGVEEINELLYKLNFFDILLAKLNDV 195
DB 200 QQAKKHLQ--EEFDAS-----LEEKDQYISVLQTVSLKQLRNG 238
QY 196 CANDYCOIPFNKIRANELDVLKKL-----VFGYR-----KPLDN 230
DB 239 -----PMN-----VDVLKPLPOLEPQAEVFTKEENPESGDPVVDGTSVKLTET 283
QY 231 IKDNGKMDYIKKKKTNIENINE-----LIESKTKIDKNKATKEP-EK-KKLYQAQYDL 285
DB 284 LQQRVKRQENLLKCKETIOSHKEQCTLTSEKALQEQDLERLQLEKIKDLHMAEKT 343
QY 286 SIYNKQLEEAHNLISVLE-----KRIDLK-KNEIKELDKLINEKNPPAN 332
DB 344 LI--TQLRDAKLNLEOLEQODKGWIAETKQRMHETLEMKEEIAQURSIKQM-----TT 396
QY 333 SGNTPTNLLDKNKK--IEEHEKEIKEIAKT-----IKFNIDSLF-----TDPLELE 376
DB 397 QGEELREQEKSERAFAFELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEERISLQ 456
QY 377 YYL-REKNKNIDISAK-----VETKESTEPNPNVNGVTYPLS 412
DB 457 QELSRVKQEVVDVMMKSSSEQIAKLQKHEKLARKEQELTKKLOTREREFQOMKVALE 516
QY 413 YNDINNALNELNSFGDLINFPDYTKEPSKNIYTDNERKKFNEIKEIKIEKKI--ESD 470
DB 517 -----KSQSEYLIKISQEKQOESLAELELQKAILTESE 552
QY 471 KKSIEDRSKSLNDITKEYEKLNLNIEYDSKFNNNIDITNFEKMMGKRYSYKVEKLTH---- 526
DB 553 NK-----LRDLOQEAITYRILELE-----SSLEKSLOENKNSKDLAVHLEAE 597
QY 527 ----HNTFASYENSKHLEKLTALKYMEDYSLRNIVVEKELKYKNLISKIENIEITLVE 593
DB 598 KKNHKEITVMVEKHKTE--LESCLKHQD--ALWTEKLQVLKQOYQTEMEKLEKCEQEKE 654
QY 584 NIKKDEEQLEFKKITDENKPDKEILEVDIVKQVQKVLNKNKIDELAKTQ-----L 636
DB 655 TLTKDKEIIFQAH-----EEMNEKTLEKIDVQKOTELLES--LSSELSEVLKARHKLLEELS 708
QY 637 ILKNV--ELKHNHVPNSYKQENKQEPYVILVLKKEID-----KLKVPMPKVESLINEE 688
DB 709 VLKQDTDKMKQLEAKMDEQKNHQQOVDSIKEHEVSIQTEKALKDQINLELLKER 768
QY 689 KKNIKT-EGOSDNSEPS-----TEGEITGQATTPKGQAGSALE-GDSVQAQAQKQAQPP 743
DB 769 DKHLKEHQAHVENLEADIKRSEGL-----QOASAKLDVFSQYQSAATHQTKAY-- 817
QY 744 VPVPVEAKAQVPTPPAPVNNKNTENVSKLDYLEKLYEFLNTSYICHKYILVSHSTWNEKI 803

Db 818 -----EEQLAQIQ--QKLLD-LET-----ERI 836
Qy 804 LKQYKITEEESKLSGCDPLDLFFIONNIPVMSFDSLNLSQ----LPMIYVEKEMV 860
Db 837 LTKQVAEVEAKQKDVCTELD-AHKIQ--VDLMQOLEKQSEMEQKYSQTQVYES--- 890
Qy 861 CNLYKLKDNK----IKNLLEAKKVSSTKTLSSSQPISLTPQDPEVSANDTSH- 915
Db 891 ----KLEDGNKQEQTKQILVEKENNMLQMRGQKKEIILT-----QKLSAKEDSIH 940
Qy 916 ----STNLNLSKLFPENILSG-----KNNIYQELIGOKSEN-----FY 952
Db 941 LNEEYETFKQKQKMEKVKQAKEMQETLKKKLLDQEAQKLENTALELSQKQKOFN 1000
Qy 953 EKIL---KDSDFYNESTFNFKSKADDINSINDESKKLEEDINKLKKTL-QLSFDLY 1008
Db 1001 AKLMAQANAGISDVSRLTNQEQIESLT-EVHRRLNDVSIWEKKLNQQAELQ 1059
Qy 1009 NKYKLKLERLFDKKTGKVMQI-----KLTLLKEQLESKLSLNINPKHVLIQ 1057
Db 1060 ETHEIQLQ--EKEQVLAELKQILLFCGEKEEMKKEITWLKEGVKQDITLNELOEQLK 1116
Qy 1058 NFSVFN-----KKKEAIAET--ENT-----LENTKILKHVKGLVYNYNG 1097
Db 1117 OKSAHVNSLAODETKLKAHLEKLEVDLNKSLKENTFLQEQVLVELKMLAEEDKRVKSEL-- 1174
Qy 1098 ESSPLKTLSE--SIQT--EDNYASLEN-----FKVLSK-----LEGK 1133
Db 1175 -TSKLTITDEEFQSLKSHKSNKSLDEKSLFEFKLSBELAQLDICKCKTEALAEKTN 1233
Qy 1134 DNLNLEKKKLSYLSGLHH-----LAEKLEKVIKKNVNTGNSPS 1172
Db 1234 ELINSSSKTNAILSRISHCOHRTKVKKALLIKTCTVSELEAQLRQITBQON-TLNISF 1292
Qy 1173 ENNT-----DVNNALBSYKFLPEGDVATVVSSESGSDTLEQSQPKKPASTHVGAESNTIT 1228
Db 1293 QOATHOLEEKENQIKSMK-----ADIESLVTEKEALQEGNQQAASE---KESCITQ 1343
Qy 1229 TSONVDDREVDVIV-PIFGESEEDYDILQGVVTGEATVPSVIDNLSKIENEYEVLYK 1287
Db 1344 LKKEUSEINAVTLMKEELKKEKVEISSKQTLNLVQ---LQNSISLSEKEAAISLR 1400
Qy 1288 PLAGYRSLKQLENNV--MFFNVN-----KDLNRSFRNRE 1323
Db 1401 K--QYDEKCELLDQVDSFKVDTLSEKISALEQVDDWSNKFSEWKKKAQSRFTQHQ 1457
Qy 1324 NFKNVLESDLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSYNYIKDSIDTDFINPANDVL 1383
Db 1458 NTVRELQIQL---ELKSKEAYEKDEQINLLKEELDQCKRFDCLKGEMDD----- 1505
Qy 1384 GYKILSEKYSKSDLS-----IKKYINDKQGENEKY 1414
Db 1506 ---KSKMEKESNLETTELKSTARIMELEDHITOKTIFIESLNEVLKNYNOOKDIEHEL 1562
Qy 1415 LPFLNNITLYKTNKDKIDFVILHLEAKVL---NVTYE-KSNVEYKIKELNY----- 1462
Db 1563 VQKLQHFQEL---GEKDNVRKAEAEKILTLENOVYSKAELETKKELEHVNLSVKSK 1618
Qy 1463 ---LKTID-----KLADFCK 1475
Db 1619 BEELKALDRLESESAKLAELKR 1642

RESULT 19

Y109_YEAST
ID Y109_YEAST STANDARD; PRT; 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 195.1 kDa protein in DNA43-UB11 intergenic region.
GN Y11149C.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule T., Moulé T., Odeil C., Pearson D.,
RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DSC-1994) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z38059; CAA86129.1; -;
DR PIR; S48385; S48385.
DR SGD; S0001411; MLP2.
KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;

Query Match 5.0%; Score 423; DB 1; Length 1679;
Best Local Similarity 19.7%; Pred. No. 4.6e-07;
Matches 335; Conservative 296; Mismatches 583; Indels 488; Gaps 78;

Qy 198 NDYCOIPNLKIRANELDVLKLVGYRK--PLDNKIKNVGKME---DYIKKN-KKTEN 251
Db 6 SEFLNVPEE-SLQGVTPVPLKRL--YKKAIFERSEEVTKLVNLDVDEIKSYSRISK 61
Qy 252 INELIEESKKTIDKNKATKEE-----EKKLYQAQYDLSIYNNKOLEBAHNLISVL 302
Db 62 LKQLLDES---SEQNTAKEELNGLKQDLNEERSRYREID--ALKQLHVSHEAMREV 115
Qy 303 --EKRI-----DTLKNENIKELDKINEIKNPPAN----- 332
Db 116 NDEKRVKBEYDIWQRDQNDLSNDDLKKNLRLKRLKLMENILQRCKSNAISLQLKYD 175
Qy 333 -SGVTPNTLLDKNKIEE-----HEKEKEIAKTKFNIDSFTDPL 373
Db 176 TSVOEKELMLQSKLLIEEKLSFSKKTLEETVKSSHVENLEEKLYQMSNYEFT--- 232
Qy 374 ELEYLRKKNKIDISAKVETK-----ESTEPNEYPNGVTYPLSYNDINNA--- 419
Db 233 ---YKFLNKNQKLSQSVSEKVKEMKNLKDTSVEKAEFFSKWPTLQNMNDLLRSQLS 289
Qy 420 -----LNELNS-----FGDLNPFDTK---EPSKN----- 442
Db 290 LEKDCSLRAIEKNDNSCRNPEDTVIDELDTKLRLKSKNECORLQNVMDCTKEEA 349
Qy 443 -----IYTDNE--RKKFINEIKKIEKK-----IESDKK-----SYEDRS 478
Db 350 TMTTSVAVSPTVYKGLFSQIKVLKRLKQIKERNQKQFQNLQLEDFILELHKTPELISFKERT 409
Qy 479 KSLNDITKEYEKLNEIYDSKFNNDILT-----NFEKMMGRYSY-KVEKLTTH 526
Db 410 KSLHEHLKRSTELLETSVLTTRKQREITSRLQKINGEANIHSVKORLQAROVKLL 469
Qy 527 HNTFASYN-SKHNLEKLTALKYMEDYSLRN-----IVVEKELKYKKNLSKIENET 580
Db 470 LNTSAIOETASPLSQDELISRLKILESSNIVNENDSQAIITRLVFPNSVNLQEKNYEL 529
Qy 581 L-----VENIKKDEEQLFEKKITKDNKPDKEKILEVSDI--VKVOQVKVLLMKNKD 629
Db 530 LNCIRILADKLENVEGKQDKTLQKVENOTIEAKDAITELNINAKMETRINILLRDS 589

```
QY 630 -----ELKKTOL-----ILKNVELKHNHVPNS 652
Db 590 YKLLASTEENKANTSVTSMEAREKKIRELEAEELSSSTKVENSAILONLRKELLYYKKQ 649
QY 653 YQENKQEPYLLI-----VLKKEIDKLKVPFK-----VESLINEEKKNIKTE-GQS 698
Db 650 CKKKTTFLEDFENFKGLAKEKERMLEAIDHLKAELEKQKSWPSYIHVEKERASTELSOS 709
QY 699 DNSEPTEGEI-----TGATTKPGQAGSALGSDVQAQAOBQAOQPPVPV 746
Db 710 RKIKSLEVEISKLKETASFTPTKESLTRDPEQC-----CKEKKELQ-----M 753
QY 747 VPVEAKAQVPTPPAPVNNKTNVSKLD-----YLEKLYEFNTSYICHYILVSHSTM 799
Db 754 RLKSEI-----SHNENKMFSSKEGOYKAKIKLENNL-----787
QY 800 NEKILQYKITKEESKLSLSCDPLDLFLNIONNIPWYSMFDLSLNNLSQLPWEIYKEM 859
Db 788 -EKLSDLSQKIQEIESIRKCDKSLKW-AQNTI-----DDTENMKSLLTETLSNKET 838
QY 860 VCNLYKLKNDKIKNLEEAKKVSTSVKTL-----SSSSMOPLSLTPQDKPEVSANDDTSH 915
Db 839 TIE-----KLSSLENDLKELRTKFOYKELQNSDASTLEPTLRKELEIQIQLKD---- 890
QY 916 STNLNLSKLFENILSLGNKNIIYQELIGQ--KSSNFYEKI-LKSDTFYNESFTNFVK 972
Db 891 ---ANSQIQAYEIIIS--SNEALIELKNELAKTNEYDAKIELEKKEKWAREDLRLR 945
QY 973 SKADDINSINDESKR-----KLEEDINKLKTQLSPDLYNKYKIKLERL-----1018
Db 946 GEIGETRALQPKLKEGALHFVQOSEKLRNEVERIQKIE-----KIEKMTTIVQ 994
QY 1019 FOKKKTGVGYKMOIKK--LTLKQELESKLSLNNPKHVLQNFVFFNKKKEAEAEIETE 1075
Db 995 LCKKEMSQYSTMKENKDLSELVIRLEDAADC-----QAEITKTK 1036
QY 1076 NTLENTKILL-KHYKGLVYNGESSPLKTLSEESTQTEDNVASLENFVKLSKLEKLD 1134
Db 1037 SSLYSAQDLDKHER---KWMEEKADYERELISNIEQTES--LRVENSVLIEKVDVTAAN 1091
QY 1135 NLNLEKKLSYSSGLHHLIAELK-----EVIKNKNTGSPSNNTDNNNALES 1184
Db 1092 NGDKHLKLVLSFNLRHENSLETKLTCTCKRELAFVKOKN---DSLEKTIINDLQRTQL 1148
QY 1185 YKFLPEGTDVATVWSESGSDTLEQSKPKKPASTHYGAESNTI--TTSQNVDDVDDVII 1242
Db 1149 SEK---EYQCSAVIIDEFKDITKEVTQ-----VNILKENNALQKSLKNVTEKNE--I 1197
QY 1243 VPIGSEBEDYDLGQ--VVTGEAVTPSVIDNLSKIENIEYVLYLKLPLAGVYRSLKKQL 1300
Db 1198 YKQLNDRQBEISRLDRLQITKEQV--SINSNKILVYSEMEQCKQR-----YQDLSSQQ 1250
QY 1301 ENNVMTFNVVKDI-----LNSRFKNRFKNVLESDDLIPYKDLTSSNVVVKDPYK 1351
Db 1251 K-----DAQKDIEKLTNEISDLKGLSSAENANADLENKF-----NKLKQOAEH 1295
QY 1352 FLNKKERDAF-----LSSYNIKDSITDITNFANDVLGYKILSEYKSDLSQSIKKYIND 1406
Db 1296 KLDASKKQAALTNELNELKAIKDLQDLHFEN-----AKVIDDTKLKAHQLESDVYS 1350
QY 1407 KOGENEKYLPLFNNIETL-----YKTVNDKIDLFIHLEAKVNLNTYKESN-VEVYKIKE 1459
Db 1351 RDHEKDYRTLMEEIESLAKELQIFKTANSSDAF-----EKLKVNKEKEKDRIIDERTKE 1406
QY 1460 LNYLKTIDQKLADFKNNNFVGLADIJSTDYNHNNLLTKFLSTGCVFENLAKTVLSNLLDG 1519
Db 1407 --FEKLQETL-----NKSTSSEAYSKD-----IETLKKEWLKEYEDE 1443
QY 1520 NLOGNLISQHCVKQCPONSGCFRHLDERECKLLNFKQK-GDKCVENPNP-TCNEN 1577
Db 1444 TLRRIEAEENLKKRLRLSEERIQIISKRKE-----ELEEEFRKLEKENSAGSLFLDN 1498
QY 1578 NG-GCDADAKCTEEDSGSNGKK 1598
```

```
Db 1499 KSGGDAEEELWNPSKGNSE 1520
RESULT 20
MYSI_YEAST
ID AC P08964;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-I isoform (Type II myosin).
GN MYO1 OR YHR023W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=91088308; PubMed=2263482;
RA Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;
RT "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide
sequence.";
RL Nucleic Acids Res. 18:7147-7147(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Willcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";
RL Science 265:2077-2082(1994).
RN [3]
RP SEQUENCE OF 1-760 FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=88111539; PubMed=3322809;
RA Watts F.Z., Shields G., Orr E.;
RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
division.";
RL EMBO J. 6:3499-3505(1987).
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; X53947; CAA37894.1; -
CC EMBL; X06187; CAA29550.1; -
CC EMBL; U10399; AAB68872.1; -
CC PIR; S05806; S05806.
CC PIR; S12323; S12323.
CC PIR; S46773; S46773.
CC HSP; P08799; LMDN.
CC SGD; S0001065; MYO1.
CC InterPro; IPR000048; IQ-region.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
```


Db 1130 IKGDIDSLFGE-----NDMLKTAKESNEAEIDRLKOKLQRSIENAKKYS 1173
Qy 1013 LKLERL---FDKKTGVGKYM-QIKKLTLLKEOLESKL-----N 1047
Db 1174 DALDKLRPEYDRLOQLYREKIKQAEINTQAVODLESRLNQSRRELDAATDKLIASEGDRN 1233
Qy 1048 SLNPNPKHVLQFNSVFFNKK-----KEAEIAETENTLENTKILLKHKYGLVKY-YN 1096
Db 1234 ALRSEVEKIQHEVOFMREOLLKRTDVEYOAAALSDLVNAHRTAEDGRVNAVQALEARKYEIN 1293
Qy 1097 GESPLKTLSESTQEDENYASLEN-----FKVLKSLKGLKDKNMLNLEKKLSVLSGLH 1151
Db 1294 DLQSRDLNAEQLYTLQOQYVAVENEDMLYDALRHSMDTRVTINRFIL-----GVD 1348
Qy 1152 HLIAELKVIKKNKVTGNSPSENNTDNNALSYKKFLPECTGVATVVSSESGSDTLEQSO 1211
Db 1349 ESMDEKKTIVLOTOKSPDGKSKERFDISDL-----DTNQLKLGRIKLELEKNE 1398
Qy 1212 -----PKKPASTHVGAE-ESNTITTSQNVDDVDVIVIFGESEEDYDDLQ-QVVT 1261
Db 1399 YRDALDRKKKKSIESHIKQETIET-NIEDQLVDV-----EERKRTLEMLRLAS 1447
Qy 1262 GEAVTPSVIDNI-----LSKIENEY-EVLVLYKPLAGVYRSLSKKQLENNVMVT 1306
Db 1448 AKQLRSQEEALKORDEERSHMKLKIAPFEMEARGEKQALRQLNELVRLNLRKDLTAQGD 1507
Qy 1307 FNV-----NVKD-----ILNSRFKNFKNVLESDDLIPYKDLT 1340
Db 1508 LGVLHDHERYVYAKFHLESKLKQESQEQIRLLVANFETERNLSNLEKVRDLASRLQOT 1567
Qy 1341 -SSNVYVVKDPYKFLNK-----EKRDKFLSYNYIKDSIDT---DINPAN 1380
Db 1568 ESKNADMKEDNRLKKDLKASTNEAELRRTIDQNSRVVSDNQILKQLESQAQNDLSNAN 1627
Qy 1381 D-----VLGYKILSEKYSKDSLDST---KKYINDKQGENEKYLPFLNNIETLYK 1426
Db 1628 NRKQOLENELLVSELRLKQFSDNANRIIDLQRLHLDTAENDKKRKTNRNLSLE---K 1684
Qy 1427 TVNPKIDLVHLEAKVLNVTYKSNVEVKI-----LNYLKTIQDLADFKNNNFVG 1481
Db 1685 TVS-----QORTIETIRQQLSLALNERNTLONDLRLQRR----- 1720
Qy 1482 IADLSTVDYNNHNLTKFLSTGWTFENLAKTVLSNLLDGNLQGMNLISQHCVKKQCPONS 1541
Db 1721 LARMETEKINN--DKYDELEKIRASLIKRI--ELLDEKRTMENILHETALQREAISS 1776
RESULT 22
MYS2_DICDI STANDARD: PRT; 2116 AA.
AC P08799;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Levinand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
Dictyostelium discoideum";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
RA Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a

phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
Dictyostelium myosin heavy chain";
FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
discoideum complexed with MgADP.Befx and MgADP.ALf4-";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
truncated head of Dictyostelium discoideum myosin to 2.7-A
resolution";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
Dictyostelium discoideum myosin motor domain to 1.9-A resolution";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammAS, and MgAMPPNP complexes
of the Dictyostelium discoideum myosin motor domain";
RL Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization
of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
Dictyostelium discoideum myosin motor domain";
RL J. Mol. Biol. 274:394-407(1997).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
(MLC-2)
CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CORTEX.
CC -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
COPIES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CVS AT THE SH-1
POSITION (688).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

Qy 1529 OHQCVKQCPQNSGFRHLDERECKCLNLYKQEGDKCVENPNPTCNENGGCDADAKCT 1588
 Db 1908 -RETVEAEADSKS-----EAEQSKRLVELEEDAR-----RNQKEIDAKEI 1948
 Qy 1589 EEDSGSN 1595
 Db 1949 AEDAKSN 1955
 RESULT 23
 ALMI_SCHPO
 ID ALMI_SCHPO STANDARD; PRT; 1727 AA.
 AC Q9UTK5; Q13313; Q9UT8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-JUN-2001 (Rel. 40, Last sequence update)
 DE Abnormal long morphology protein 1 (Sp8).
 GN ALMI OR SPAC1486.04C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN
 RC SEQUENCE FROM N.A.
 RP STRAIN=972;
 RX MEDLINE=2184801; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cetrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 RN [2]
 RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=972;
 RX MEDLINE=20123449; PubMed=10660053;
 RA Jimenez M., Petit T., Gancedo C., Goday C.;
 RT "The almi1 gene from Schizosaccharomyces pombe encodes a coiled-coil
 protein that associates with the medial region during mitosis."
 RL Mol. Gen. Genet. 262:921-930(2000).
 RN [3]
 RP SEQUENCE OF 644-834 FROM N.A.
 RC STRAIN=968 h90;
 RX MEDLINE=20223868; PubMed=10759889;
 RA Ding D.Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hiraoka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 fission yeast cells by the use of a GFP-fusion genomic DNA library."
 RL Genes Cells 5:169-190(2000).
 CC -!- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
 CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN

CC CYTOKINESIS.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL133357; CAB62414.1; -
 CC EMBL; AF010473; AAB65416.1; ALT_INIT.
 CC EMBL; AB028012; BAA87316.1; -
 CC Coiled coil.
 CC DOMAIN 57 361 COILED COIL (POTENTIAL).
 CC DOMAIN 443 463 COILED COIL (POTENTIAL).
 CC DOMAIN 542 740 COILED COIL (POTENTIAL).
 CC DOMAIN 804 1106 COILED COIL (POTENTIAL).
 CC DOMAIN 1223 1427 COILED COIL (POTENTIAL).
 CC DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 CC DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 CC SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
 Query Match 4.7%; Score 399.5; DB 1; Length 1727;
 Best Local Similarity 20.6%; Pred. No. 2.7e-06;
 Matches 336; Conservative 279; Mismatches 562; Indels 453; Gaps 80;
 Qy 231 IKDNVGMKEDYIKRKKTKTINELIE-----ESKKTIDKKNATKE-----E 273
 Db 6 LEDDQLVHEFL---DVSFEDIKPLVSVNGFAVTSIAIKTKVKDINALKDQLVQEVNHE 62
 Qy 274 EK-----KKLYQAQVDLSYKNQKLEEAHNLISVLSEKRTDITLKK-ENIKELDKINEIKN 327
 Db 63 HKENVLTKKINLEQLOQSSNNQAESESNLISVLSEKRTDITLKK-ENIKELDKINEIKN 117
 Qy 328 PPPANGTNPNTLDDK-NKITEEHEKEKEIAKTIKFNIDSLFTDPLEYLYLRKKNKI 386
 Db 118 -----TTNQSLRRANSELQEQSKIASQLSIAKDQIEAL-----QENNS 157
 Qy 387 DISAKVETKESTEPENYNGVTYPLSYNDINNALNELNSFGDLINPFDYTKEPSNIYTD 446
 Db 158 HLGEOVQS-----AHQALSDI----- 173
 Qy 447 NERKK-----FINEIKKIKIEKK-----KIESDKKSYEDRSKSL 481
 Db 174 PERKQHFASSSSRVKEILLVQEKASVLSLQSDHSKVCCKELEVSSROVQDEKKL 233
 Qy 482 NDITKEYEKLKNE---IYDSKFNNDITLNFPEKMMG-----KRYSYKVEKLTHHTTFAS 532
 Db 234 AGLAQONTE-LNEKIQLFQEKRSNYSSDGNISKILETDPTSIKELEEEVE---TKRLTAL 290
 Qy 533 YENSKHNLEKTKAL--KYMEDYSLRNIWE-----KELYYKNLISKIENEIEFTLVEN 584
 Db 291 WESKSELQSEVAALQEKLTQQSLYNNVTEELNNKQOLISENSRLQEKYDSVYSE 350
 Qy 585 IKKDEEQ-----LFE---KKITDKNKPEKILEV-SDIVKVOVKVLLANKNK 630
 Db 351 LQVKNENKNTSVSACVGLFSPLAQKLSAVON-PESFTKTVYSDNNKLOQKVSLLKLQDR 409
 Qy 631 L-----KKTOLILKNVE---LKNHNVHPNS-----YKQENK 658
 Db 410 LTNKFSSECFCEQVKORIPVVKQORSEIVRNINYMFLSESLTSSNNLTQVQAEALLTKMR 469
 Qy 659 QEPYYL-----IVLKEIDKLKVPMPK-----VESLNEEKKNKTTEGO 697
 Db 470 QEACYQLQTASRTQCSDLSEREVICLMAELDLHNETKSRNVFATVQVQALDEYAQNESTA-- 527
 Qy 698 SDNEPSPTEGTEITGAATKPG-----QAAGSALLEGDSVOAQAQE-OKAQPPVPVPVP 749
 Db 528 ---SETLVNKELANFSSIKREAVSKTLEIREKVRALCED-VEIQKQTVQVQISNAVKNEN 583
 Qy 750 EAKAQVTPPPAVN-NKTENSVKLDYLEKLYEFLNTSYICHKYILVSHSTMNEKILKQYK 808

```
Db 584 TLSEQIKNLESELNKKIKNELNRLNKLKMLATF-----RSSILSHNSSAGNI----- 634
QY 809 IYKEESKLSDDPDLLENQNNIPVMYSMPDLSLNNLSQ-----LWEIYKEKWCN 862
Db 635 -----DOKKSID--ESTRELEKNEYEVYRNEMTAQIESKRNQDLSEMAIRKLENS 687
QY 863 LYKLG-DNOKIKNLLKEAKVSTSVKTLSS--SSMOPLSLTPQDKPEVSANDDTSHSNL 919
Db 688 KYQQQLSTDRLTNANDVFAFKKEAKELASINQNLQDI-ISRODQRAKFAEELLHVNSL 746
QY 920 NNSLKLFENILSGK---NKNYIYQELIGQSSNFYEKIL-----KSDTFYNES 966
Db 747 AERLK-GELNASKGEXDLKRTQERLI-----SEN--DKLLAERLMSLSVSLQTLNQ- 798
QY 967 FTNFVSKADDINSLNDESKRKKLEEDINKLKTQ-LSFQDLNKKYKLLERLFDKXIV 1025
Db 799 -----QOLSDAARKVFESKESLSLSQKLKESNKMNDLHSLQK-----SLEKSG 846
QY 1026 GYKMKQIKKLTLLKEOL-----ESKLSLNNPKHV-LQNFVSF---FNKKK 1067
Db 847 IEYSSRIKTLMLKESLSEDNRRKLLDNQOMMEIKLQELNGVIELEKQRFSTLEAKFTQK 906
QY 1068 EBIATENTLEN--TKILLKHVGLVYNGESSPLKTYLSESIOTEDNYA--SLENPK 1123
Db 907 NTSYSREALLSSLSLQSKH-----TSLESQYNSLRNIEQLQ 946
QY 1124 VLSKL-----EGKLDKNLLEKKLSYLSLSSGLHHLIAELKEVKN-----KN 1165
Db 947 AASKLAEMVVRKTYDEYRTLSSEKKNHLKITS--LEQRIVLQDEIASSSLRUCN 1004
QY 1166 YTGNSPS-----ENNTDVNNALSY---KKFLPBGTDV--ATVVSSESGSDTLBQSOP 1212
Db 1005 ITKDSERVALLEENKHLNLSHRNAEKQHLKENDYKQOOLLVTDLRKTRDYB- 1063
QY 1213 KRPASTHVAES-----NTITTSQNDVEDDVIIPIFGESE-----EYDDLQGV 1259
Db 1064 -KELLRHADARSTLQKREDYTKALEQVEDLNKEITAKAGINESQPPFISEKEDPLRQEV 1122
QY 1260 VTGEAVTPSVIDNLSKIENEVEVLKLP-LAGVYRSLKQLENNVMYTNVNVKDLNLSR 1318
Db 1123 YVLKKQNALMLTQOLSSNLFABEITSPSDLSVMKGLSLQNHVVKRISKEMEIISCOR 1182
QY 1319 ----FNKREPNKVNLESLIPYKDLTSSNVVVKDPYKFLNKRKDKFLSSYNYIKDSIDT 1375
Db 1183 QLLFLENKKLKTVE-----SSNRVIAD-----LQRGITEKDVSS---TSESVGER 1225
QY 1376 INFANDVLGYKTLSEKYKSDLSIKKYINDROGENEKYLPPL-NNIETLYKTVNDKIDL 1434
Db 1226 SNYLNWV----ALLNESKSLRENLER-----NEEVITELREKIELT-----KTDL 1267
QY 1435 FVTHLEAKVLNITYEKSNEYVKIK-ELNVLKTIQDKLADFKKNNNFVGI---ADLSTDYN 1490
Db 1268 -----ANFLNKEQLESQLOTEKAQVKKLENSNEYKRRHQEILLSLNSSTSSD 1318
QY 1491 HNNLLKFLSTGMVFENLAKTVLSNLDGNLQGLMNI--SQHQVYKQCPQNSGCFRHLID 1548
Db 1319 ASRLKNELVSKNLEELAQEI-----GHLKSELETVKSKSDLENERAQNSKIEQLE 1372
QY 1549 ERE---ECKLLNKKQEGDKVENPNPTCNENNGGDDADAKTEEDSGSGNGKKITC---- 1601
Db 1373 LKNTKLAARWTKYEQVNVKSLKHN-----QIRQOLSQRTSELEAKVAECHOLN 1422
QY 1602 -ECTKPDSPV 1610
Db 1423 EQLNKPSATP 1432

RESULT 24
RA50_YEAST
ID RA50_YEAST STANDARD; PRT: 1312 AA.
AC P12753;
DT 01-OCT-1989 (Rel. 12, Created)
```

```
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN DNA repair protein RAD50 (153 kDa protein).
OS RAD50 OR YNL250W OR N0872.
OC Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RE821;
RC MEDLINE=89276917; PubMed=2659437;
RA Alani E., Subbiah S., Kleckner N.;
RT "The yeast RAD50 gene encodes a predicted 153-kD protein containing a
RT purine nucleotide-binding domain and two large heptad-repeat
RT regions.";
RL Genetics 122:47-57(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RC MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldener U., Beinhauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT from the left arm of chromosome XIV from Saccharomycetes cerevisiae.";
RL Yeast 13:849-860(1997).
CC -!- FUNCTION: INVOLVED IN DNA DOUBLE-STRAND BREAK REPAIR (DSBR). THE
CC RAD50/MRE11 COMPLEX POSSESSES SINGLE-STRAND ENDONUCLEASE ACTIVITY
CC AND ATP-DEPENDENT DOUBLE-STRAND-SPECIFIC EXONUCLEASE ACTIVITY.
CC RAD50 PROVIDES AN ATP-DEPENDENT CONTROL OF MRE11 BY UNWINDING
CC AND/OR REPOSITIONING DNA ENDS INTO THE MRE11 ACTIVE SITE.
CC -!- SUBUNIT: FORMS A COMPLEX WITH MRE11.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: X14814; CAA32919.1; -
DR EMBL; X96722; CAA65494.1; -
DR EMBL; Z71526; CAA96157.1; -
DR PIR; S05808; BWBYDL.
DR SGD; S0005194; RAD50.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR004584; Rad50.
DR TIGRFAMS; TIGR00606; rad50.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Meiosis.
FT NP_BIND 34 41
FT DOMAIN 185 347
FT DOMAIN 403 558
FT DOMAIN 617 672
FT DOMAIN 734 1108
SQ SEQUENCE 1312 AA; 152568 MW; 58A0AA173AC5677E CRC64;

Query Match 4.7%; Score 393; DB 1; Length 1312;
Best Local Similarity 19.8%; Pred. No. 3.3e-06;
Matches 264; Conservative 251; Mismatches 443; Indels 376; Gaps 60;

QY 204 PFNLKIRANELDVLKLVGYRKPLDNINKNYKQLEEAHNLISVLEKRIDTL-KNENIKELLDKI 263
Db 168 PSNLKKKDFEIQAMK---FTKALDNLKSIKKDMSVDIKLLKQSVHLK-----L 214
QY 264 DRKNKATKEEKKKLYQAYQDLSIYKQLEEAHNLISVLEKRIDTL-KNENIKELLDKI 322
Db 215 DKDRS---KAMKLNTHQLQTKIDQYNEEYSEIESQLNEITEKSKDLFKSNQDFQKILSV 271
QY 323 NEIKNPPPPANS-----GNTPTNLLDKNKKTEEHEKEI---K 355
Db 272 ENLKNTKLSISQVQRKLSNIDILDSKPDQLNLANFVKVLMKNNQLRDLETDISSUK 331
```



```
Query Match      4.6%; Score 386.5; DB 1; Length 2748;
Best Local Similarity 20.6%; Pred. No. 1.2e-05;
Matches 377; Conservative 300; Mismatches 640; Indels 513; Gaps 96;

QY 53 LNEGTSGTAVTSTPGSKGSVASGGGSGVASGVSAS-----GGSVASGG 98
DQ 28 LSGEYSTNGVTRMANKLKADECGSGDEGDKTRFSISILSKRETKDVLPEFAGSSSHG 87
QY 99 SVASGSGSRRRT-NPSDN-----SSDSDAKSYADLKHRVNRVLLTIKELKYPQLPD 149
DQ 88 VLTANSKDMNFTLEUSENLVECRKLQSSNEAKN-----EQIKS-LKQIKESLSDKIEE 141
QY 150 LTNHMLTCLDNIHGFYKYLIDGEEIN-ELLYKL-NFYFDLLRAK-LNDVCANDYCOIPEN 206
DQ 142 LTNQK-----KSPMKELDSTKDLNWDLESKLTNLSECRQLKELKKYKTEKSWNDEKES 194
QY 207 LKTRANELDVLKLVFG-----YRKPLDNIKNVGMEDYIKNNKKTIEINEL 255
DQ 195 LKLLKTDLLEILTITKNGMENDLSSQKLYDKETSELKE---RILDNNENDRLIISVSD 251
QY 256 IEB-----SKKTIDKNK---NATKEEEKKKLYQAY---DLSIYN-----K 290
DQ 252 TBSINLSQNRTERIKIQKQDADAKASISLARKVKYQKYOQHTSDTTVTSDPDSEGT 311
QY 291 QLEEAHNLISVLEKRTIDTLKKNENIKELDKINEKNPPANGSNTPNTLLDNK---K 347
DQ 312 SEEDIFDVIDIEMTETGSEVEDISEDLVKYSEKNMILLNSDYSKYNLQKSESASRP 371
QY 348 EBEKEIKEIKIATIKENIDSLTDP--LELEYLRKKNIDISA-KVETKESTEP-NEY 403
DQ 372 KODELTKVAENL--NMIALNDNDNYSKFEFSLESHIKYLEASGYKVLPLEEFENLNE- 428
QY 404 PNGVTYPLSVNDINNALNELNFGDLINPFDYTKPKSKNLYTDNERK-----KFINE 455
DQ 429 --SLNP-SYNLYKEKQLQAKIPIQDOSTFNLKKEPTIDFLPLTSKIDCLLIPTKDYND 485
QY 456 IKKIK-----IEKK-----KIESDKKSYE 475
DQ 486 LFESVKNPSTEQMKKLEAKDLQSNICKWLEBERGCKWLSNDLYFSWVKNKIETPSKY- 544
QY 476 DRKSLSNDITKEYELL--NEYDSEFNND-----LTNFEKMGKRYKYV 521
DQ 545 -----LSDRAKEQDVLIDTKALEGLKNPTIDFLREKASADYLLKKEIDYVSPSLEYLV 599
QY 522 E--KLPHHTF--ASYEN---SKHNLEKLTALKYMEDYS--LRNIVBEKELKYKNLIS 572
DQ 600 ERAKATNHLLSDSAYEDLVCKENPD-----WEFLKESAKLGHVTVSNEA--YSELEK 552
QY 573 KIEN-BEITLVENIKKDEBQLFE---KKITKDENKDEKILEVSDIVKVOQVILLMNK 627
DQ 653 KLEQPSLEYLVEHAKATNHLLSDSAYEDLVCKENPDMEFLKE-----696
QY 628 IDELKKTLQILKVELKHNHVPNSYKQENKQEPYLYLVKKEIDKLVFMPKVESLINE 687
DQ 697 -----KSAKLGHVTVSNEAYSELQK-----YSELEKEVEQ-----PSLAYLVEH 736
QY 688 EKKN-----IKTEGQSD--NSEPSTEGETTGQATTKPGQAGSALGDSVQAQAO--EQOK 739
DQ 737 AKATDHLLSDSAYEDLVCKENPDVEFLKESAKLGHVTVSSEYSELQRYSELEKEV 796
QY 740 AOPVPVPVPEAKQVTPPPAPVNNKTVNSKLDYLEKLYEFINTSYICHKYLIVSHSTM 799
DQ 797 EQPSLAYLVEHAKA-----TDHLLSDSAYEELVKCKENPDMEFLKESAKLGHVTV 848
QY 800 N-----EKILKQYKI-----TKEEKSLSCDPLDLLFNIIQNNIPVWYSHFDSLNS 846
DQ 849 SNEAYSELEKLEQPSLAYLVEHAKATDHLLSDSAYEDLVCKENSDVEF-----LKEK 903
QY 847 LSOLFMEIYEKEMVCNLYKLNKDKIKNLEAKKVTSTVKTLSSSQMPLSITPDQKPE 906
DQ 904 SAKLGHVTVSNEAYSELEKLEQPSLAYLVEHAK--ATDHLLSDSAYEDL-----952
```

```
QY 907 VSANDTSTSHNTLNNSLKLFFENILSGKNKNYIQEL-----IGQKSSNFYE-----KIL 956
DQ 953 VKCKENPDMEFLKESAKLGHVTVS-----NEAYSELEKLEQPSLEYLVEHAKATNHLL 1008
QY 957 KDSDFYNESFTNFVYSKAD-DINSLNDESKRKKLEEDINKLKKTLQLSDFDLYNKYKLL 1015
DQ 1009 SDS-----AYEDLVCKENPDMEFLKESA-----KLGHVTV-VSNEAYSELEK 1052
QY 1016 ER-----LFDKKTKGVKYMQ---IKKTLKLEQLESKLSNNPK--HVLQNFVSFFNK 1065
DQ 1053 EOPSLEYLVEHAKATNHLLSDSAYEELVKCKENPDVEFLKESAKLGHVTVSNEAYSEL 1112
QY 1066 KKEAEIAETENTLENTKILLKH-----YKGLVYKYGESSPLKTLSESIOTEDNYASL 1119
DQ 1113 EKKLEQPSLEYLVEHAKATNHLLSDSAYEELVKC--KENPDVEFLKESAKLGHVTVS 1170
QY 1120 ENFKVLSKLEGLKDNLNEKKLSYL-----SSGLHLLAE--LKEVINKNYTGNSPSE 1173
DQ 1171 EAYSELEK-----KLEQPSLAYLVEHAKATDHLLSDSAYEDLVCKC-----1212
QY 1174 NNTDVNNALSEYKFLPECTDVA--TVVSESGSDTLEQ--SOPK-----KPASTHYG 1221
DQ 1213 ENPDV-----EFLKESAKLGHVTVSNEAYSELEKLEQPSLAYLVEHAKATDHLL 1264
QY 1222 ABS-----NTITTSQNVDE-----VDDVIIVPIGSESEEDYDDL-----1256
DQ 1265 SDSAYEDLVCKENPDMEFLKESAKLGHVTV-----SNEAYSELEKLEQPSLEYLVE 1318
QY 1257 -GOVVTGEAVTSDVIDNLSKIENEYEVLYLAKPLAG-----VYRSLKKQLENNVM 1305
DQ 1319 HAKATNHLLSDSAYEDLVCKENPD--DMEFLKESAKLGHVTVVSNKEYSELEKLEQPS 1377
QY 1306 TF-----NVNVDIILSNFRNKREN-----FKNVLES-----1331
DQ 1378 EYLVKHAEOIQSKIISIDFNTLANPMSMEDMASKLOKLEVQIVSNDEYIATKNTMEKPDV 1437
QY 1332 DLIPYK-----DLTSSNVYVKD---PYKFLNKEKROK---FLSSYNYI-----1368
DQ 1438 ELRLSKLGYHIIDTITYNELVSNFNSPTLKFIEEAKSGYRLIEPNEYDLNRIATTP 1497
QY 1369 -KDSITDINFANDVLGYKILSEKY---KSDLS--IKKYINDKQGENEKYLPFLNNIET 1423
DQ 1498 SKEED---NFKQI--GCYALDSKEYERLKNLSLENSKRFIE-----ENALDLDLVLDKT 1549
QY 1424 LYKTVDND---KIDLFVHLEAKVLNY-----TYEK-----SNVEYKIKELNY 1462
DQ 1550 EQYAKMDNASKSLIP---STKVLDFTVPAPQASAEKSSLOKRTLSDIENELKALGY 1606
QY 1463 LKTIOQKLAADFKNNNFVGIADLSTDYNNHLLTF-----LSTGMVFENLAK--TVLSNL 1516
DQ 1607 VAIRKENLPNLEK-----PIVDNASKNDVLNLSKPSLVPLST--EYDNNRKEHTKILNI 1660
QY 1517 L-DGNLQ-----GMLNISQHCQVKKO 1536
DQ 1661 LGDPSIDFLKECKEYQYMLIISKHDYEEKO 1690

RESULT 26
HMW2_MYCPN
ID HMW2_MYCPN STANDARD: PRT: 1818 AA.
AC P75471;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cyathderence high molecular weight protein 2 (Cyathderence accessory
DE protein 2).
GN HMW2 OR MPN310 OR MP526.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
```

RE MEDLINE-97105885; PubMed-8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkil E., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE-97252497; PubMed-9098066;
RA Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
RA Hermann R.;
RT "Transposon mutagenesis reinforces the correlation between Mycoplasma
RL pneumoniae cytoskeletal protein HMW2 and cytoadherence";
RL J. Bacteriol. 179:2668-2677(1997).
CC -I- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC IN THE MYOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A000051; AAB96174.1; -
DR EMBL; U59896; AAB52527.1; -
DR PhosSite; P75471; -
KW Cytoadherence; Structural protein; Coiled coil; Complete proteome.
FT DOMAIN 31 880 COILED COIL (POTENTIAL).
FT DOMAIN 919 1607 COILED COIL (POTENTIAL).
FT DOMAIN 1644 1755 COILED COIL (POTENTIAL).
FT DOMAIN 1786 1817 COILED COIL (POTENTIAL).
SQ SEQUENCE 1818 AA; 215622 MW; 66DF4B08F0FCB0 CRC64;
Query Match 4.58; Score 378; DB 1; Length 1818;
Best Local Similarity 19.68; Pred. No. 1.5e-05;
Matches 334; Conservative 308; Mismatches 595; Indels 470; Gaps 80;
QY 115 DNSSDSDAKS---YADLKHVRVRLTLIKELKYPQ-----LFDLTNHLMLTCDN 160
DB 32 ESAANDAQVVELQTQLLAEIKKLENEIKALKAESQPDPHNNARQSLASLNRLVNE 91
QY 161 IHGPK----YLIDGYERINELLYKLNPFYDLR-----AKLNDVCANDYQCIPFNKLI 209
DB 92 YNNEFQKNYVDRVAELNN---KARFQDELKRLQENAAFNRYAN-WADFQSNVQL 147
QY 210 RANELDVLKLVGYRPLDNKDNVGMEDYIKNNKTTINENE-----LIEESKKT 262
DB 148 KLDQFQAL-----IDQNOTIKQLNEQIAANOGLIDQNVQR 183
QY 263 IDKNKNAATKEBEKKLYQAQYDLISYKQLEEAHNLISVLEKRIDTLKKNENIKELLDKI 322
DB 184 LQNHSLDQOERDALLYVDH---LYN-ELYELN-----QKRLVGIEYEATYQDLYSAD 234
QY 323 NEIKNPPANGNTPTNLLDNKKKIEBEKEIKEIAKTIFNIDSLFTDPLEYLYIREK 382
DB 235 AELQNVYETIAQNANE-----QKQDAYWAQKQVEQIQ----- 270
QY 383 KNKIDISAKVETKESTENPENGVTYPLSYND-----INNALNELNFGDLINPFDY-TK 437
DB 271 -----TTKQELVDEE-----STLKVRLNDADFYINSRLAELDLSKINERDFVSK 316
QY 438 EPSKNIVTDNRKRFINEIKIEKIKIEKIESDKKSYED-RSKSLNDITK-EYEKLL--- 492
DB 317 EQADVV-----KASLANLTK-----EKERSAEKDSFERLNTALNDINRMWENALPAK 366
QY 493 ---NEIYDSKFNNNIDLTFEE---KMMGKRY-SYKVEKLTHTHTTFASYENSKHNLEK--- 542

DB 367 HLEQOQVEFERKQOESLLKLETEHKQKRGEPKIESEAKSEALLIQE---RELLERRE 424
QY 543 ----LTKA-LKYMEDYSLRNIVVEKE---LKYKNLI---SKIENEIETLVENIKKDEEQ 591
DB 425 IDLLTQASLEYEQORRTNQVLKLEKROVQOQHFONHAKKKLDOKRHYLAEQKRIDEEQ 484
QY 592 LFEKKITKDNKPKDEKILEVSDIVKVOQVYLLMNKIDELKKTOLILKNVLEKKNHVPN 651
DB 485 IFK---LKEKIATERRELEKLYLVKKQKQ-----DQKENDLLIFEKQLR---QYQA 529
QY 652 SYKOENKQEPYLLIVLKEIDKLVKPMKVESLINEEKKNIKTEGQ---SDNSEPSTEGEI 709
DB 530 DFENEIEEKQNELFASOKLSQKSFQTLKNKEAELNQAKTAEDWAHLKONKHHADLEI 589
QY 710 TQQTATPQGOAGSALGSDSVQAQO-----EOKQAPPPVPVPEAKAVPT 757
DB 590 FLEGEFNHLOQEKHL-----LEARTQFDNRVSLLSARFKQKQAE-----LVKQKSLQ 639
QY 758 PPAPVNNKTENV-----SKDYLEKLYEFLNTSYICHKYILVSHSTMN---EKILQYKIT 810
DB 640 LTAAFNKEQEAVERDKDRLANLEKQKEMLGDK-VHQF---DENSLNISKLAERELAI 694
QY 811 KEESKLSKCDPLDLLFNIONNIPVMYSMFDLSNLSQLPMEI-YEKEMVNCNLYKLKDN 869
DB 695 KFEKELEAAQKQLSLDN---NNAGLKLQDLKLSLTERLELEASKERILDFYD-ESS 751
QY 870 DKIKNLEEAKKYSTVSTKTLSSSSMQPLSTPQDKPEVSANDDTSHST---NLANSKLKF 926
DB 752 RIADYESDLOARLAEVKTL-----KNOQETAASKERELKVALEKLNQAKRAF 800
QY 927 -----ENILSGKNKNIYQELIGQSSSENFYEFKILKSDTFYNESFTNFVSKADDINSL 981
DB 801 LQIRKQOLLEIASVK---QOLAQKA-----NLLKNQQAELDKQ 835
QY 982 NDESKRKKLEEDINK---LKTTLQSLFDLYNLYKULKLER---LFDKKT---VGKYM 1031
DB 836 TEELEAAFLQDQTKLEKALH---SVKSKQEL-LERERSFLQKQREFAHVAHVGKFRQ 891
QY 1032 IKKTLTLLEQLESKLSNPNPKHVLFQSV---FFNKKEAEIAETENTLTKLKH 1088
DB 892 VHFKTQMQRL-SEFNQOQSEQIKRETELKIAFADUKDYQLFELQKNOBQFQLEQKH 950
QY 1089 K-GLVYKYNIGESSPKLTLSIESIQTDENYASLENFKVLSKL-----EGKLDN----- 1135
DB 951 ELELLAQKAE-----LKQELEQKATASQDQDTVOAKLDLARQOHELELRQAFNAQ 1004
QY 1136 ---LNLEKKKLSYSSGLHHLIAELKE-----VTKNNYTGNSPSENNTD---VNNAL 1185
DB 1005 SLSLNKQREQLTNQVKVLH---GELKKRHEKLTILKDLAEKEKDKHKKDAEIN---QRF 1058
QY 1186 KFLPEGTQV-----ATWVSESGSDTLQEQSQPKKPASTHVGA 1223
DB 1059 KOFENEYADFQAKKRELQELNQTIRNLEQSNASLLKKRNQTLTDFALLRK-----VQ 1111
QY 1224 SNTITTSQNVDDVDVVIPIFGESEEDYDGLQVVTGEAVTFSVIDNLSKINENEVEY 1283
DB 1112 HNTQTRVQLNTQIKEFL-----EKNFKQKASDEAALQKALLIKRLRSFASKLQOREA 1166
QY 1284 LYLPLAGVYRSKLQLENNVMTFNVVNVDILNSRFNKRNF----- 1325
DB 1167 LAIQLEFDRDEQKSEIN-----NAKLOEQFKLEKQNFDEAKQKLIJEFKQCCORL 1220
QY 1326 ---KNVLESDLIPYKDLTSSNYVVKD-----PYKFLN----- 1354
DB 1221 DVERKLLKQLVLQNLNLSKSYLTTKNRADLSQQQLQHKYANLLEKLEKQAKRALDKKH 1280
QY 1355 -----KEKRDPLSSYNIKD-----SIDTDI 1376
DB 1281 RAIYKMAQFVSELROEKKOLLSSAQKQVDDKSRLLQEQNRHLQNLSSFTKKKROSLEHDI 1340
QY 1377 NFANDVLGYKILSEKYKSDLSIKKYINDKQGENEYKLPPLNNIET-----LYKTV 1428

Db 1341 N-----KFDQRRKAVSILNSHKK-LKQKEGELQGLQKLSKKTKQIEQFESKLYQO- 1392
QY 1429 NDKID-----LFVHLEAKVLNTYEKSNVEVKIKELNVLTKIQ---DKLADFKKNNN-- 1478
Db 1393 REKLDRQRTLSKLHRELKAQNEATAHKNREVLEIENYKKELQRLTTEKSEFDNKNRL 1452
QY 1479 ---FVGIADSLTDYNNHLLTKFLSTGMVFENLAK---TVLSNLDGNGLNQGLMLNTSQHQC 1532
Db 1453 FEYFRKI-----RNETEKEAHIKTVLEETQKRHLVETEAVKLHLQKQSIISKGE 1504
QY 1533 VKQCPQNSGCFRHLID-BREECKCLLN 1558
Db 1505 LKEIKERSVDRISHTNKKOREELNSLLH 1531

RESULT 27

RA50_METJA STANDARD; PRT: 1005 AA.
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 AtFase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Scott J.R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodex A.,
RA Utterback T.R., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U67572; AAB99331.1; -
CC TIGR; MJ1322; -
DR InterPro; IPR003439; ABC transportr.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N.1.
DR Pfam; PF02483; SMC_C.1.
DR ProDom; PD000006; ABC transportr; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).
SEQUENCE 1005 AA; 119387 MW; 9BBB48173E78F3 CRC64;

Query Match 4.4%; Score 372; DB 1; Length 1005;
Best Local Similarity 20.9%; Pred. NO. 1.2e-05;
Matches 272; Conservative 204; Mismatches 364; Indels 464; Gaps 67;
QY 363 ENIDSLFTDP-----LELEYLYREKNNKIDISAKVETKESTEPNE--YPNGVTYPLSYND 415
Db 56 FNYDPIITTKGKSVVVELDFEVNGNNYKI-----IREYDSGRGGAKLYNGKPYATTISA 110
QY 416 INNALNEL-----NSFGDLINPFDYTKPEPSKNIYTDNERKKFIN-EIKEKIKIEKKIES 469
Db 111 VNKAVNEILGVDRNFMNSI-----YIKQ-----GEIAKFLSLKPEKLETVAKLGI 158
QY 470 DKKSIEDRSKSLNDITKYEKLLNIEYDSKFNNIDILNFKMMGKRYSY---KVEKLT 526
Db 159 DE--FEKCYQKMEIVKEVKEKRLRI--EGELNYK---ENYEKELKNKSQLSEKKNKLM 212
QY 527 HNTFASYENSKHNLEKLTALKYMEDYSLRNIVVEKELKYKKNLISKIENETIETLVENIK 586
Db 213 IN-----DKLNKIKKEFED-----TEKLFNEW----- 235
QY 587 KDEOLFEEKITKDENK-----PDEKILEVSDIVKVOVKVLLMNK-----ID 629
Db 236 -NKKLLYEKFINKLEERKALELKNQELKILEYDLNTVVEARETLNRHKDEYKYSLV 294
QY 630 ELKLTQILILKNVELKHNHVPNSYKOEKQEPYIYLVLKKEIDKLKVFMPK----- 680
Db 295 EIRKTESRLR--ELKSHY---EDYLLKTKQ-----LEIITGDIKLEKFINFSKYRDDIDN 345
QY 681 VESLINEEKKNIKTEGQSDNSEPSTEGETGOATTKPGQOAGSALGDSVQAOAEOKA 740
Db 346 LDTLLNKIKDEIE-----RVETIKDLLEBK-- 371
QY 741 QPPVPVPVPEAKAQQVTPPAPVNNKTNVSKLDYLEKLYEFLNTSYICHYILVSHSTMN 800
Db 372 -----NLENELEKLEKRYCECKEY--EKL-----ELE 401
QY 801 EKILKOYKITPE-----BESKLSSCDPLD-----LFNIQNNIPWYMSFOSLNSLSQ 849
Db 402 EKAVERNKLITLEYITLLQEKKSIEKNINDLETRINKLLEETKNIDI-----ESIENSLK- 455
QY 850 LPWEIYEXEMWC-NLYKLLK-----DNDIKNLLEEAKKVSTSVKTLSSSSMOP 896
Db 456 ---EIEEEKKVLLENQEKIELNKGLGEINSEIKRLKILDELKEVEGKC-----P 503
QY 897 LSLTPQD---KPEVSANDDTSHSTNLNLSKLFENILSLGKNKIYQELIGOKSENFYE 953
Db 504 LCKTIDENKMKELI---NQHTQLNNK-----YTEL----- 532
QY 954 KILKSDTFYNESFTNFVKSADDDINSUNDESKRKLLEEDINKLKTQLSLDLYNKYL 1013
Db 533 -----EIN-----KKIREIKDEIKLKEIDKEENLTKLTP 565
QY 1014 KLERLFDKKTKYVKMOIKLTLKLEQESKLSLNINPKHVLNFSVFFNKKKEAIEAE 1073
Db 566 YLEK-----QSQLEELKELKNYKEQLEDEINK---ISNVI---NKPVDEI-- 607
QY 1074 TENTLENTKILLKHYKLVKYNGESSPLKLSIESIOTEDNYASLENFKVLSKLEGLK 1133
Db 608 ----LEDIKSQLNKFNFQVLSAVSYLSNVDEEGIR--NRIKEIENI-VSGWNKEKCR 660
QY 1134 DNLNL---EKKKLSYLSGLHHLIAELKEVTKNKNYTGNSPENNTDVNNALESYKFLP 1190
Db 661 EELNKLREDEIREINLKOKNLNKNKELEIE-----IENRRSLK--FDKYKEYL- 708
QY 1191 EGTDTVAVVVSSEGSTLEQSQKPKPASTHVGAESNTITTSQNVDDDEVDDVIIVPIFGESE 1250
Db 709 -----GLTEKLEE-----LKNIKDGL-----E 725
QY 1251 EDYDOLGQVVTGEAVTPSPVIDNLSKIENEYEVLYLKLPLAGVYRSLSKOLENNVTFN 1310
Db 726 EIYNICNSKIL-----AIDNKKRYNKEDIEIY-----LNNKILEVNEK 764
QY - 1311 VKDIILNSRPNKRNFKNVLESDLIPYKDLTSSNYYVVKDPYKFLNKEKDKFLSSYNIKD 1370

Db 765 INDI-----EERISVIN-QKLDEINYN-----EEHKKIRELY-----ENKROELDNVRQKT 811
QY 1371 SIDTDINFANDVLGYKILSEKYSDDLSDIKKYINDKGENEKYLPFLNNIETLYKTVND 1430
Db 812 EIEGTG-----EYLKDVESLKARLKE-----MSNLE-----KEKE 842
QY 1431 KIDLFVHLEAKLVNYYEKSNSVVKIKELNYLKTIO-----DKLADFKNKNFVGIADLS 1486
Db 843 KLTFEVEYLDK--VRRIFGRNGFOAYLRE-KYVPLIQKYLNEAFSEFDLPYSFV---ELT 896
QY 1487 TDYN-----HNNLLT-KFLSTGWFVFNLA-----KTVLSNLLDGNL-----QGMNLNIS 1528
Db 897 KDFEVRHAPNGVLTIDNLGG---EQIYAVALSRLAIALNIGNVECIILDEPTVYLD 953
QY 1529 QHOCVK-----KOCPONSGCFRHLDERECKLLNLYKOECD 1564
Db 954 ENRRAKLAEFRKVKSPQMIITTHRELEADVADVIINVRKDG 997
RESULT 28
AKA9_HUMAN STANDARD; PRT; 3911 AA.
AC Q99996; Q9U0Q4; Q9UQH3; Q9Y6T2; O14869; O43355; O94895; Q9Y6B8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
DE (PRKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
DE protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
DE (hyperion protein) (Yotiao protein) (Centrosome- and golgi-localized
DE PRN-associated protein) (CG-NAP).
GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=98151389; PubMed=9482789;
RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;
RT "Yotiao, a novel protein of neuromuscular junction and brain that
RT interacts with specific splice variants of NMDA receptor subunit
RT NR1.";
RL J. Neurosci. 18:2017-2027(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RX MEDLINE=99219864; PubMed=10202149;
RA Witczak O., Skallhegg B.S., Kerker G., Bornens M., Tasken K.,
RA Jahnson T., Oerstavik S.;
RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring
RT protein located in the centrosome, AKAP450.";
RL EMBO J. 18:1858-1868(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99287934; PubMed=10358086;
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that
RT anchors multiple signaling enzymes to centrosome and the golgi
RT apparatus.";
RL J. Biol. Chem. 274:17267-17274(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Kemmer W.A., Weiss S., Schwarz U.;
RT "Cloning of Hyperion.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RC TISSUE=Gastric parietal cell;
RX MEDLINE=99115654; PubMed=9915845;
RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RA Trotter K.W., Milgram S.L., Goldenring J.R.;
RT "AKAP350, a multiply spliced protein kinase A-anchoring protein
RT associated with centrosomes.";
RL J. Biol. Chem. 274:3055-3066(1999).
RN [6]
RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RC TISSUE=Lymphoblast;
RA Hinds K., Sutterer C., Becker M., Hawkins M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RC TISSUE=Lung;
RA Milgram S.L., Goldenring J.R., Schmidt P.H.;
RT "AKAP350: A multiply spliced family of proteins with centrosomal
RT association.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RL for large proteins in vitro.";
RN DNA Res. 5:277-286(1998).
RN [9]
RP SEQUENCE OF 17-1800 FROM N.A.
RA Wu X., Graves T., Bradshaw H.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND
CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL
CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN
CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-
CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR
CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS
CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.
CC -1- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N
CC (PKN). PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)
CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.
CC -1- SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND
CC CYTOPLASMIC IN PARIETAL CELLS.
CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3/CG-
CC NAP, 4/YOTIAO, 5 AND 6/AKAP350; ARE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY
CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.
CC -1- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC -1- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.
CC -1- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR
CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AJ131693; CAB40713.1; -
DR EMBL; AB019691; BAA78718.1; -
DR EMBL; AJ010770; CAA09361.1; -
DR EMBL; AF026245; AAB86384.1; -
DR EMBL; AF083037; AAD22767.1; -
DR EMBL; AF004013; AAB96867.1; ALT_FRAME.
DR EMBL; AF091711; AAD39719.1; -
DR EMBL; AB018346; BAA34523.1; -
DR EMBL; AC000066; AAC60380.1; ALT_FRAME.

DR Genew; HGNC:379; AKAP9.
KW Coiled coil; Alternative splicing; Polymorphism.
FT DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.
FT DOMAIN 164 314 COILED COIL (POTENTIAL).
FT DOMAIN 944 1022 COILED COIL (POTENTIAL).
FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).
FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).
FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).
FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).
FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).
FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).
FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
FT DOMAIN 3587 3689 POLY-LEU.
FT DOMAIN 3726 3730 GLN-RICH.
FT DOMAIN 203 292 GLN-RICH.
FT DOMAIN 321 1010 GLU-RICH.
FT DOMAIN 1846 2772 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPLIC 17 28 QLOBE1 -> LATRD (IN ISOFORM 4).
FT VARSPLIC 1637 1642 MISSING (IN ISOFORM 4).
FT VARSPLIC 1643 3911 MISSING (IN ISOFORM 3).
FT VARSPLIC 2175 2182 SADTFQKVE -> Q (IN ISOFORM 6).
FT VARSPLIC 2175 2183 VGFYVMCFSTLC -> GSSIPELAHSDAYOTREICSS
(IN ISOFORM 2, ISOFORM 3 AND ISOFORM 5).
FT VARSPLIC 2895 2907 MISSING (IN ISOFORM 5).
FT VARSPLIC 2895 2948 STTQFHAGMRR -> ALSLTSMQHHRSARPTAPLFFETLSH
(IN ISOFORM 6).
FT VARSPLIC 3901 3911 SLG (IN ISOFORM 6).
FT VARIANT 1347 K -> QO.
FT CONFLICT 76 76 /FTID-VAR_010926.
FT CONFLICT 475 475 E -> Q (IN REF. 3).
FT CONFLICT 554 554 M -> I (IN REF. 3).
FT CONFLICT 638 638 E -> G (IN REF. 3).
FT CONFLICT 663 663 R -> S (IN REF. 3).
FT CONFLICT 913 913 N -> S (IN REF. 3).
FT CONFLICT 956 956 H -> N (IN REF. 3).
FT CONFLICT 980 982 K -> N (IN REF. 3).
FT CONFLICT 997 997 QKH -> PKP (IN REF. 1 AND 2).
FT CONFLICT 1001 1001 Q -> P (IN REF. 1 AND 2).
FT CONFLICT 1020 1020 Q -> P (IN REF. 1 AND 2).
FT CONFLICT 1028 1028 N -> D (IN REF. 3).
FT CONFLICT 1626 1626 V -> E (IN REF. 3).
FT CONFLICT 1703 1703 R -> P (IN REF. 1 AND 2).
FT CONFLICT 1707 1707 V -> T (IN REF. 3).
FT CONFLICT 1707 1707 V -> G (IN REF. 3).
FT CONFLICT 1802 1803 MISSING (IN REF. 5).
FT CONFLICT 1843 1843 A -> P (IN REF. 3).
FT CONFLICT 1956 1956 I -> V (IN REF. 3).
FT CONFLICT 2027 2027 V -> D (IN REF. 5).
FT CONFLICT 2157 2158 EI -> HE (IN REF. 7).
FT CONFLICT 2169 2169 E -> V (IN REF. 3).
FT CONFLICT 2314 2314 L -> R (IN REF. 3).
FT CONFLICT 2851 2851 I -> N (IN REF. 8).
FT CONFLICT 2957 2957 E -> D (IN REF. 3).
FT CONFLICT 2983 2983 P -> S (IN REF. 3, 7 AND 8).
FT CONFLICT 3087 3087 Q -> H (IN REF. 3).
FT CONFLICT 3218 3218 O -> H (IN REF. 3).
FT CONFLICT 3307 3309 ESE -> QSQ (IN REF. 3).
FT CONFLICT 3751 3751 P -> A (IN REF. 3).
FT CONFLICT 3833 3833 T -> S (IN REF. 3).
SQ SEQUENCE 3911 AA; 453664 MW; 3FB1CB1C819B47AA CRC64;
Query Match 4.3%; Score 365; DB 1; Length 3911;
Best local Similarity 18.3%; Pred. No. 8.9e-05;
Matches 285; Conservative 302; Mismatches 560; Indels 412; Gaps 69;
QY 173 E1NELLYKLNFFY-----DLRLAKNDVCANDYQIQIPNLKIRANELDVLKLVFGYRK 226
DB 242 ELTEQSQKLIQIQLOQASSETLRNSTHSTADLLQAKQOILTHQOOLEEQDHLLDYQK 301

QY 227 PLDNKIDNVGKMDYIK-----KNKKTIIENINELIEESKTKIDKNKKNATKEEKKKLYQAO 282
DB 302 KQEDFTMOISFLQEKIKVYEMQDKKVENSKKEEOEKETIEELNTKIIIEEKKTLTK 361
QY 283 YDLSIYNQLEEAHNLISVLEKRIDTLK-----KNEIKELLDKINEIKNPP 330
DB 362 DKLTATADLLGELQIQVOKNOEIKNMKLELTNSKOKEROSSEETKOLMGTVTEELQK--- 418
QY 331 ANSGNTPTLLDNKK-----IEEHEKEIKEAKTIKFNIDSIFTDPL-ELEYLRE 381
DB 419 -----RNKDSOFETDIIVORMQETQORLEQLRAELDEMGQOIVQMQLR 466
QY 382 KKNKIDISAKVETKESTEPNEYPNGVTYPLSYNDI-----NNALNELN-SFGDIL 430
DB 467 QH-----MAQMEEMKTRHKGEMENALR---SYSNITVNEQIKLMNVAINELNIKLODTN 518
QY 431 NPFDYTKEPSKNIYTDN-ERKFFINEIKIEKKIESDKKSYEDRSKSNLDTKE-- 487
DB 519 SOKEKLKEELGLILEKCALQLEDVLEELSFQIQARQRTIAEQESKLEAHKSL 578
QY 488 -YEKLLNNEYDS-----KFNNDIDLTFE---KMMGRYSYKVEKLTHTHTFASYNS 536
DB 579 TVEDLKAELIVSASESKKELELKAHEAVNTYKIKLEMEKKNVLDLDRMA-----ESQ 630
QY 537 KHNLEKTKAL-----KYMEDYSLRNIYVEKELK-----YKKNLSKIEIENEIETLV 582
DB 631 EAELERLRTOLLFSHEEELSKLDELEIGHRINIEKIDNLGIHYKQKQDGLQNMESQKI 690
QY 583 ENIKDEEQLFEK-----KITK-----DENRPEKILEVSDIVK-----VQVOK 621
DB 691 ETMQFEKDNLTQKQNLQILEISKLDLQOOSLVNSKSEMTLQINLEQKEIETLRQEK 750
QY 622 VLLMKNKIDELK-KTOLILKNVELKHN-----IHPNSYKQENKQEPYIIVLKKED 672
DB 751 GTLEQEVQELQKTELLEKQMEKENDLOEKFAQLAEANSILKDEK-----KTLED 801
QY 673 KLVKVPKVESLINEKKNIKTEGSDNSEPT--EGEITQOATTKPQQAGSALGEGS 729
DB 802 MLKIHTP-----VSQERLIFLDSIKSKSDSWEKEIEIIEENEDLKQOC----- 848
QY 730 VQAOAQEOKQ-----AOPPPVPVPPEAKQ-----VPTPPAPVNNKTYENSKLDY--- 774
DB 849 IOLNEEIEKQRTTFAEKNFVNYQEOLEEYACLLKVKDDLEDKSKNOE---LYKSK 904
QY 775 LEKLYEFLNTSYICHYILVSHSTMNE-----KILQYKITKEE 814
DB 905 LKALNEELHLQRIPTTVKMKSSVFDEKDTFVAETLEMGEVVEKDTTLMEXLEVTKREK 964
QY 815 SKLSS--CDPLDLFNQNNIPVMSFDSLNSLSOLFMEIYKEMVCLYKLD--- 868
DB 965 LELSQRSLDLSEQLKQKHGEISFLNEEVKSLKOEQVSLRCRELEIIINHNRANVQSC 1024
QY 869 NDKIKNLE---BAKKVYSTSVKTLSSSSMOPLSITPODKPEVSANDDTSHSTLNLSL 923
DB 1025 DTQVSSLLDGVTMTSRGAEGSVKVKNSFGGESKIWVEDK--VSPENMTVEESKQEQ 1082
QY 924 KLFINLSLGKNKNIYQELIGOKSSNFYKILKQSDTFYNES-----FTNF 970
DB 1083 -ILDHLPSVTKESS---LRATOPSEN--DKLOKELNVLKSEQNDLRLQMEAQICLSLV 1135
QY 971 VSKKADDINSNDESKRKKL-----EEDINKLAKTLQLSFDLYKYLKLERLF 1019
DB 1136 YSTHVDQVREYMEKDKALCSLKEELIPAEQEKIKELQKIHOLEQ-----TMKTQETG 1190
QY 1020 DKKK---TVGKYKMOI-KKLTLLKEQESKLSLNNP-----KHVLQNFVSFFNKKK 1067
DB 1191 DECKPLHLLIGLKQAVSECSYFIQTLCVSGEYTPALKCEVNAEDKENSODYISENE 1250
QY 1068 EAEIA---ETENTLNTKILL---KHYKGLV-----KYNGESSPLK-TLSESTIQ 1111
DB 1251 DPQLQDYRYEVQDFQENMHMTLLNKVTEYENKLLVLQTLRLSKTWGQQTGDMKLEFGEENLP 1310
QY 1112 TED-----NYASLENPKVLKLEGKLDNLNLEKKKLSYLSGLHHLIAELKEVIN 1163

```
Db 1311 KBETFLSIHSQNTLEIDVNHK--SKLSSLDQLEKTKLEQVOELESLSLQQQLKE 1368
Qy 1164 KNYTSGNSPENTDYNNALESYKKFLPEGTDVATVVSSESGDTLQSQPKKPASTHVGA 1223
Db 1369 -----TEQNYAE--IHCLQKRL-----QAVSES-----TVPSLPVDSVWITE 1405
Qy 1224 SNTITP-----SONVDEVDVIVIPFGSESEYDGLQGVVTGEAVTPSVIDNLSK 1276
Db 1406 SDAQRTMYPGSCVKKNIQDGTIE--FSGEFGVKEE-----TNIVKL 1443
Qy 1277 TENEYEVLYLPLAGVYSLKQLENNVMTFNVNFKDILNFRFNKRENFKNVLESDLIPY 1336
Db 1444 LEKOYQ-----EQLEEVAKVIVS-----MSTAFQAQOTELSRISGG----- 1479
Qy 1337 KDLTSSVYVVPYKFLNKRKDKFLSSYNYTKDSIDTIDINFANDVLGY--YKILSEYK 1394
Db 1480 KENTASS---KQAHAVCOEQ-----HYF-----NEMKLSQDQIGFQFETVDVFEK 1523
Qy 1395 SLDLSIKKYINDKQGENKYLPLFNN-----IETLYKTVNDKIDLVFVHLEAKV 1443
Db 1524 BEFKPLSKEL-----GEHGKEILLSNDPHDIPESKDCVLTITISEEMFSKDKTFIVR----- 1574
Qy 1444 LNYTEKSNVEVKIKELNYLXTI---QDKLADFKNNFVGTADLSTDYNNHNLTKFL 1499
Db 1575 -----QSIHDEISVSSMDASRLMLNEQLEDMRO-----ELVRQYQEHQQAPELL 1620

RESULT 29
DMD_CHICK
ID DMD_CHICK STANDARD; PRT; 3660 AA.
AC P11533;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dystrophin.
GN DMD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098331; PubMed=3062582;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA.";
RL Nucleic Acids Res. 16:11815-11815(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Muscle;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal coding and 3' untranslated regions between man and chicken.";
RL EMBO J. 7:4157-4162(1988).
CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE PLASMA MEMBRANE.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN, ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.
CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
```

```
CC -----X13369; CAA31746.1; -
DR PIR: S02041; S02041.
DR HSSP: P46939; 18HD.
DR InterPro: IPR001589; Actbind_actnin.
DR InterPro: IPR001715; Calponin-like.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR002349; WW.
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR InterPro: IPR000433; Znf_Z2.
DR Pfam: PF00307; CH; 2.
DR Pfam: PF00397; WW; 1.
DR Pfam: PF00435; spectrin; 22.
DR Pfam: PF00569; Z2; 1.
DR PRINTS: PR00403; WWDOMAIN.
DR SMART: SM00033; CH; 2.
DR SMART: SM00150; SPEG; 21.
DR SMART: SM00456; WW; 1.
DR SMART: SM00291; Znf_Z2; 1.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS00021; CH; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS00020; WW_DOMAIN_2; 1.
DR PROSITE: PS01357; ZF_Z2_1; 1.
DR PROSITE: PS0135; ZF_Z2_2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; Zinc-finger.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 19 123 CH 1.
FT DOMAIN 138 241 CH 2.
FT REPEAT 341 449 SPECTRIN 1.
FT REPEAT 450 558 SPECTRIN 2.
FT REPEAT 561 669 SPECTRIN 3.
FT REPEAT 721 830 SPECTRIN 4.
FT REPEAT 832 936 SPECTRIN 5.
FT REPEAT 945 1047 SPECTRIN 6.
FT REPEAT 1050 1156 SPECTRIN 7.
FT REPEAT 1159 1265 SPECTRIN 8.
FT REPEAT 1268 1369 SPECTRIN 9.
FT REPEAT 1470 1570 SPECTRIN 10.
FT REPEAT 1573 1678 SPECTRIN 11.
FT REPEAT 1681 1782 SPECTRIN 12.
FT REPEAT 1789 1981 SPECTRIN 13.
FT REPEAT 2013 2103 SPECTRIN 14.
FT REPEAT 2106 2211 SPECTRIN 15.
FT REPEAT 2214 2321 SPECTRIN 16.
FT REPEAT 2472 2574 SPECTRIN 17.
FT REPEAT 2577 2683 SPECTRIN 18.
FT REPEAT 2686 2799 SPECTRIN 19.
FT REPEAT 2802 2904 SPECTRIN 20.
FT REPEAT 2906 2928 SPECTRIN 21.
FT REPEAT 2931 3037 SPECTRIN 22.
FT DOMAIN 3052 3085 WW.
FT ZN_FING 3304 3351 Z2-TYPE.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;

Query Match 4.3%; Score 361; DB 1; Length 3660;
Best Local Similarity 18.8%; Pred. No. 0.00011;
Matches 343; Conservative 297; Mismatches 592; Indels 588; Gaps 82;

Qy 140 KELKYPQLFDLTNHLMT-----LCDNTHG-FKYLIDGYEEN 175
Db 1212 KEVKVKLITDSYNNPIAKAPPAANEALKKELDLVLTYSQRLSGKCKTL-----EEVW 1267
Qy 176 ELLYKLVNFYDLRLAKRNDVCANDYQIPFNLIKIRANE-----LDVKKLVFG 223
Db 1268 ACWHELLSYLDAENKWLNEV-----ELKATENIQGAEISELSLERLM-- 1315
Qy 224 YRKPLDNIKDNVGMEDYIKKNKKTIENTINELIE-----SKKTIKKNKA 269
```

Db 1316 -RHPEDN-RNQIRELAQTLDGGILDELINLEKENTRWEELOQBAVRROKSLEQSIQS 1373
QY 270 TREEEKKLYQAOYDLSYNNKOL-----BEAHNLISVL---EKRIDTLKK- 311
Db 1374 AOETD-KTLRLTOESLAADKOLATAYADRVDAQAQVPOBAQKIQSELTSHEISLEEMKKR 1432
QY 312 ---NENIKELLKINIKNPPANSNTPTLLDKKKTEEHEKEIKEIAKTIKFNIDSJ 368
Db 1433 NRKESAKRVLSQIDVAOK--KLQDYSMKFRFLFOKPAFEOQLQOECKRILDEVKLVQPKL 1490
QY 369 FTDPLE-----LEYLREKKNKIDISAKVET-----KESTE-PNEYPNGVT-Y 409
Db 1491 ETKSVQEVQVSHLDHCKMLKYLKSLSEVKSEVETVKTGRQIVQKQOOTENPKELDERLTAL 1550
QY 410 PLSYNDINNAL-----NELNSFGDLINPED----- 434
Db 1551 KQYINELGAKVTEKKQOELEKCKLKLKRLKEINSLFEWLAATDVLTTKRSVQGMPSNLD 1610
QY 435 ---YTKPSKNIYTDNERKKEIETNEIKIEKIEKKIESDKSYEDRSKSLND-----ITK 486
Db 1611 AETAWGKATRKETEKRVQVLKNICDILGENLKVTLKGKES--LVEDKLSILNSNIAVTS 1667
QY 487 EYEKLN-----EYDSKFNNDILTNFEKMMGKRYKVKELTHHNTFASYENSK 537
Db 1668 RAEWLNLLMEYQKHMEAFDQKVN--VTT-----WIYRAETILLDESQKQPQKE 1716
QY 538 HNLEKLTALKVMEYD--SLRNIVVE----- 561
Db 1717 EYTLKRLKAEINMHPKVDVSRQAVDLMTNRGDHCRKVTEPKLSELNHRFAAISRIKSG 1776
QY 562 -----KELKY-----KNLSIKIENIEIETLVENIKKDEQLEPKKITKDKENPDKILEV 611
Db 1777 KPPIPLKEQEDFIQKLEPLEVEIQGV-NLKEED--FNKDMSEDESVKELLQR 1832
QY 612 SDIVKVOVKVLLMMKIDELKKTOLLKN--VELKINI--HVPNSYKQENKQEPYLYIVL 667
Db 1833 GDTLQKRI-----TDERKREEIKIKOOLLQTKHNAKLDKRSRRKKALEISHOWYQ 1884
QY 668 KKEIDKLVFMPKVESLI-----NEBKNIKTEGQSDNSEPSTEGETIGQATTKPQQOAG 722
Db 1885 KROADDLMTWLDIEKKLASLPDHKDEQKLEIGGELEKKEDLNNAVNQAEKRLSKDGA 1944
QY 723 SALEGDSVQAQAQ-----EQKQAQPPVPVPPEAKAQVTPPAVNNKNTENSVKLDYLEKL 778
Db 1945 KAVEPTLVQLSKRWDRFESKFAQ-----FRRLNVAIQ 1977
QY 779 YELNTSYCHKYLVS-----HSTMNEKILKQYKITEEESKLSS-----CDPL----- 823
Db 1978 TVLEDTTFVMTESMTVETTVVPSTYLAETLQILQALSEVEERLNSPVLOAKDCEDLLKOE 2037
QY 824 -----DLLFNIONNIPWYIS-MFDSLNNLSOLPMEIYEK-----EMVCNLYK 867
Db 2038 ECLKNIKDCGLRGHIDIIHSKKTALQASATPRETANIQDKLTQLNSQWEKVNKYRDR 2097
QY 868 DN--DKIKNLEPAKKVSVKTLVS---SSSMOPLSLTPQDKPEVSANDTSHSTNLNNS 922
Db 2098 QARFDRSK---EKWRLFHCMEKSFNEWLTETEKLS-----RAQIBAG-DVGHV----- 2142
QY 923 LKLFENILSLGKNKIYQEL---IGQ-----KSENFYEKILKDSDTFFYNESFTNEVSK 974
Db 2143 -----KTKQFLQELQDQIGRQQTWVKTANVTGEBEIBQS-----SAADANVLKQ 2187
QY 975 ADDINSINDB-----SKKKLE-----EDINKLKLKTLQLS----- 1004
Db 2188 LGNLNTRWQIECBQVEKKEKRIEENKILSEFQEDLNKILWLEETENVIATPLEPGNE 2247
QY 1005 ---FDLYNKYKLERLDFKK-----KTVGKYKMQIKKLTLLKQLESKLSNLN 1051
Db 2248 QLRDCLGKVKLRVEELLPHKILKRLNETGGTTLGASLNPER---KHKLESTLKEAS- 2302
QY 1052 PKHVLNFSVFNKKKEAEI-----AETENTLNTKILKHYKGLVKYVYNGESSP-LKTL 1105

Db 2303 -RLLKVSRLPEKQKEIBILLKQFIELNQOINQLTWTIPVKNQLELYNQVQPGAFDI 2361
QY 1106 SESIQTEDNYASLENFKVLSKLEG---KLKDNLNEKKKLSYLSS---GLHHLIAELK 1158
Db 2362 KETEAQVQAKQPNVE--EVLSK--GCHLYKEKPAHPVKKKLEDLNADWKAINHLLQLK 2417
QY 1159 E-----VIKNKNYTGNSPSENNTDNNALLESYKKFL 1189
Db 2418 EKPTFGEPALTPSGVLTSGTVAVDTOARVTKETTSFTPEMPSSVLEEV-PALADENKAW 2476
QY 1190 PGTDVAT-----VVSESGDTLEQSQPKKPASTHVGAESNT 1226
Db 2477 AELTWSLRDLREIKAQRTVGLDDINDMIKQKANMODLEQRRPOL-----DEL 2527
QY 1227 ITTSQNVDDDEVD---VVIIVPIFGESEDDL-----GOVVTGEA 1264
Db 2528 ITAAQMLKNTKSNOEARTIITDRIEKIQSOWDDVHGYLQNRROQLHEMOKDSTQWLEAQ 2587
QY 1265 VTPSVIDNLSIKIENEYEVLY-----LKPAGVYRSKLEKOLEN----- 1302
Db 2588 EABOVLEQAKAKLESWKESYTVTEALKQNSLQKQFSEIKROQMNIQEGVNDVALKPVRD 2647
QY 1303 -----NVMTFNVNKK-DILNSFRNKRENFKNVLESOLIPYKDLTSSNYYVVKDPYKF 1352
Db 2648 YSADTRKVKELMTDNINATWATINKRVSERE---AALESALLMLQEF-----Y 2692
QY 1353 LNKERDKFLLSSYNIKDSITDINFANDVLYGYKILSEKYSKSDLSIKYIINDKQGENE 1412
Db 2693 LDLEKFLAWLT-----EAETTANVLODATHKEKTLEDQPM--VRELMKQWDLQAEID 2743
QY 1413 KYLPFLANNI-----ETLYKTVNDKIDLFVHLEAKVLYNYEKSNEVVKIKELN---YLK- 1464
Db 2744 AHTDIPHNLDENGOKILRSLEGSSEDAVLLQRRLDNNFRMS-----ELRKSLNIRSHLEA 2799
QY 1465 -----TTODKLA-----DFKKNFNVGIALDSTYDHNHNLITKF---LST-GMV 1504
Db 2800 STDQWRKRLHLSQELLANLQKLEDELKQAPIG-GDIPTVOKONDVHRTFKRELKTEPV 2858
QY 1505 FENLAKTVLSNLIDGNLQGM 1524
Db 2859 INNALETVRLFLADQPVEGL 2878
RESULT 30
ID TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=14371155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U67604; AAB99663.1; -
DR TIGR: M31643; -
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR005289; GTP-binding_dom.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_N; 1.
DR Pfam: PF02483; SMC_C; 1.
DR TIGRfams: TIGR00650; MG442; 1.
KW Hypothetical protein; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 31 38 ATP (POTENTIAL).
FT DOMAIN 160 521 COILED COIL (POTENTIAL).
FT DOMAIN 673 1032 COILED COIL (POTENTIAL).
SQ SEQUENCE 1169 AA; 136634 MW; B63CE34E4C03F36 CRC64;

Query Match 4.2%; Score 353; DB 1; Length 1169;
Best Local Similarity 20.1%; Pred. No. 5.8e-05;
Matches 299; Conservative 226; Mismatches 451; Indels 510; Gaps 66;

QY 35 LEDAVLTGYSLQKEKMWLNNEGTSQAVTTTPGSGKSVASGSGSVASGSGSVASGSGV 94
DB 4 LEKIELNFKSKFKSLDIPKGF--TAI-----VGPNGSGKSNVDAILFVLGKT 51

QY 95 ASGSGSVASGSG-----NSRRTN-----PSDSSSDSAKSYADLKHVR-----NYL 136
DB 52 SAKKLRAFRFSLGIYTHNKRADFAVCILYFNENAFVNAVADKGVILRIKSSGETDIY 111

QY 137 LTIKELYPQLDFLTNHLMLTCDNIHGFYKYLIDGYBEINELLYKLNFYFDLLRAKLNVC 196
DB 112 LVWKENDKCKKMKTKH-----EIIDLFRLLGLGD-----142

QY 197 ANDYCOIPNLKIRANELDVLKLVFGYKPLDNK-----DNVGKMDYIKKKKTTIE 250
DB 143 -----NVSQGLLKIINISPIERRKIIDELSGIAEFDEKKKKABEELKKARELIE 193

QY 251 NINELTEESKKTIDKNKATKBEK-----KKLYOQYDL-----SIYKQLEEAHNLS 300
DB 194 MIDIRISEVNNLLKKLKEKEDAEKYIKLNEELKAQKAYALILKYSYLVNLENIONDI- 252

QY 301 VLEKRIIDLTKKNENIKELDKINEIKNPPANGNPNTLLDKNKKIEE-HEK-----EIKE 356
DB 253 ---KNLEELK-----NEFLSKVREI-----DVEIENIKLRLNIIINELNEKNEEVLE 297

QY 357 IAKTIKFNIDSLFTDPLELEYLYLREKKNIDIS---AKVETKESTPEPNEYPNGVTYPLS 412
DB 298 LKHSIK-----EVEIENDKKVLDSSINELKKVEV-----328

QY 413 YNDINNALNELNSFGDLINPFDTYTKPSKNITDN---ERKKFINEIKERIK---IERK 465
DB 329 --EIKENKKEI-----KETQKKIENRDSIIIEKEQKIIEBKIKNLNLYKE 373

QY 466 KIESDKKSYEDSKSLN-----DITREYKELNIEYDSKFNNNIDLTNFKKMGKRYSYKV 521
DB 374 RLKEATAESIIKHLKSEMEIADIAKNQNELYLRLKRELN-DLDN-----LINRKNFEI 428

QY 522 EKLTHNTFASNSKHNLEKLTALKYMEDYSLRNIVVEKELKYKKNLSIKIENIEFTL 581
DB 429 EK-----NMEMTKKLEETVEDVTKPLYLELE-----NLNVEIEFS 467

QY 582 VENIKKDEQLFEKTKTKE-----NKPDEKILEVSDIVKQVQVKL-----LNN 626
DB 468 KRGIREEKKKELQAKLDLHAELYKKNARIKALKKEMEELSMORAIRAILNANLPGIID 527

QY 627 KIDELKKTOLILKNVELKHNHVPNSYKOENKOEPPYILVLKKEIDKLVFMPKVESLIN 686

DB 528 IYVNLGKTK-----IEYKTAIEVAAGNRLNH-----IWVRMDDAVRAI-----KYLK 570
QY 687 EKKKNIKTQSGSDNSEPSTEGETGOATTKP-----GQOAGSALGSDSQVQAQAEQKQAO 741
DB 571 ERK-----LGRATFLPLDRIGREA-YYIDEDGVIGRA-----602
QY 742 PVPVPVPEAKAQVPTPPAPVANNKTENVSKLDYLEKL-----YEFLNTSYICHKYLIVSH 796
DB 603 -----IDLVEFDEKYRRVFEVFCNT-----VVVEN 628
QY 797 STWNEKILKOYK-----ITKEEBSKLSSCDPLDLLFIQNNIPVWYSMFSLNLSLQFLM 852
DB 629 IDIAKELAKYKRVFTLTDGVIEPISGAMIGTFFSKAKIKVDVL-----SKLNKIAD 683
QY 853 EIVEKEMVCLYKLDNDKIKNLLBEAKKVSSTVKTLSSSSMOPLSLTPODKPEVSANDD 912
DB 684 EIIAIE-----SELRIKIEIERLSKIVKSSAKKME-----715
QY 913 TSHSTNLNNSLKLFIENLSLGNKNKIYQELIGQKSSSENFYEKILKSDDTFYNESFTNFVK 972
DB 716 -----IENTLEIKKNEMKREIAEKNTIKIKELEKKNKIDL--EELSELNL 760
QY 973 SKADDINSND-ESRKKL-----EDINKLK-----TLOLSEDLNKNYKLLKLE 1016
DB 761 KREEILNRINEIESKINELIERREKTIINELKVEYSDENLKRMEIERGELKILEKAKLK 820
QY 1017 RUFDRKKTKYKMKOIKKLTLLK--EOLSKLNSLNNPKHVLFQNSVFENKKEAEIAETE 1075
DB 821 NEIDKGLTL-----VKEILPKIEELNNKVSLELNKKVILEKNISFYKESIEKNL-----870
QY 1076 NTELENTKILLKHVGLVYKNGESSPKLTISESIOETDNYASLENFK--VLSK---LEG 1130
DB 871 SILEEKR---KRYEELAK-----NLKELTEKKEOLEKEIETLERERREILRKVRDIEN 920
QY 1131 KIKDNLNLEKKKLSYSSGLHHLIAELKEVINKNKTGNSPSENNTDNNALLESYKFLP 1190
DB 921 RINE-LMWEKAYE-----SKLEE-----EERKLYLC 946
QY 1191 EGTDVATVYSESGDPTLEQSQPKKPASTHVGAESNTITTSQNVDDVDDVIIVPIFGESE 1250
DB 947 EKVDVSKELKKDIELE-----IYIGELENEIKS-----LEPVNRAI 985
QY 1251 EYDDLQGVVTGEAVTPSVIDNLSKIE--NEYEVLYLPLAGVYRSLKKQLENNYMTFN 1308
DB 986 EDYN-----YVAERYKELIEKKEVERDEKK-----YLOLMELEN-----1021
QY 1309 VNVKDLNLSRFNK--RENFKNVLESLLPYKDLTSSNYV---VKDPYK---FLNKEKRDK 1360
DB 1022 -KKKEVFMVFNKAVKNFEV-----YKEIGGIGKLSLENEKNPFEGGILIDASPRGK 1073
QY 1361 FLSSYN-----YIKDSIDTDINFA 1380
DB 1074 KLLSLDAMSGGKSLTALAFLAIOQLNPSFPYVLDEDAALDKVN 1119

RESULT 33
SBCC_CLOAB
ID SBCC_CLOAB STANDARD; PRT; 1163 AA.
AC Q97FK1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclease sbcCD subunit C.
GN SBCC OR CAC2736.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;

RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.,
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
CC -|- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (By similarity).
CC -|- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
CC -|- SIMILARITY: BELONGS TO THE SMC FAMILY. SBC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE007771; AAK80682.1; -
DR InterPro: IPR003439; ABC_transportr.
DR Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication;
KW DNA recombination; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 35 42 ATP (POTENTIAL).
FT DOMAIN 197 415 COILED COIL (POTENTIAL).
FT DOMAIN 446 1003 COILED COIL (POTENTIAL).
FT SEQUENCE 1163 AA; 135507 MW; CEF0BD2215D7A92 CRC64;
Query Match 4.2%; Score 352.5; DB 1; Length 1163;
Best Local Similarity 20.1%; Pred. No. 6e-05;
Matches 274; Conservative 212; Mismatches 448; Indels 431; Gaps 60;
Qy 204 PFNLKIRA-----NELDV-----LKKLVGYRKKPLDNTKDNVGMEDYKNNKKTIEIN 253
Db 3 PIRVRIKGLNSFENEQEFNFKELTKRGLFGFGTGTSGKTTI-----LDSITLSYG 54
Qy 254 ELIESKKTIDKNKATKEEKKLYOQYDLSIYNKOLEAHNLISVLEKRIDTLKNE 313
Db 55 EVARKSNFMNTNCS-----LNVSEFQISGKEIR-----YLVEREFRDNTG 100
Qy 314 NIKELLDKINEKPPPPANSNGTPTLLDKNKKIDEEKEKEIKAKTIKFNIDSL----- 368
Db 101 SVRSKAKIVDI-----TGDEVELEEGAKSVNEKQEI-----IGSLDDFTRTV 147
Qy 369 -----FTDPLEEYLR-----EKNKI-----DISAKVETKESTEPNYPNGVYPL-SY 413
Db 148 LPOGKFSEFLKLEGERNMLERLNFLOEYGDLSFKLARKIRK-REKENVLVGLKGY 206
Qy 414 NDIN-NALNE-----LNSFGDLINPDYTKPSKNIYTDNERKKFIN--EIKEKIKIEK- 464
Db 207 ENINEDVLKERRELLKENDFN--EASKEYLKAEETNEGEKVNGLOIEIEEKNRVRKD 264
Qy 465 ---KKIESDKKSYEDRSKSLNDITKEYEKLLENIYDSRFNNNIDLTNPFKMMGRYSYKV 521
Db 265 LMWKKDEIDLEKRAKGESSKVPY-----IDNYENILKQIDILKE 307
Qy 522 EKLTHHNTFASYENSKHNLEKLTALKYMEDYSLRNIYVVEKELKYKNLISKIENEIFTL 581
Db 308 QILSRENTMKAISLEKEDMEKKLSIAKONKEKALPKPKPKHHI-----ILDAIKE--KDL 360
Qy 582 VENIKKDEEQIFPKKTKDENKRPDEKILEVSDIVYQVQKVLMMKIDELKKTOLIKNV 641
Db 361 LDNLIKLEKRL-----QGIKELSLASN-----KEELIKQNIKDIDSLTIKIQNL 406
Qy 642 ELK--HNHVPNSYKQENKQEPYLLVLKKEIDKLKVFMPKVESLINEEKKNIKTGQSDN 700
Db 407 ESKIDNLKVPPEYKNKINEG---IFLLRNYDEKL-----HKN----- 441
Qy 701 SEPSTEGEITQATTKPGQQAAGSALSGDSVQAQAEQQAQPPVPVPEAKAQVPTPPA 760

Db 442 -----KLG-----LDCOKFQVDFEKAQSKKEML----- 464
Qy 761 PVNKNTEVNSKLDYLEKLYEFLNTSYICHKYILVSHSTMNKILKQYKITKEESKLSLC 820
Db 465 -FNKLEERSKLDYTKKLODLNKPDPKDDVLLTFOEKLNDSRQWAKYSEYNLSKAS- 522
Qy 821 DPLDLLFNQNNIPVMSFDSLNNSLSOLFMEIYEKEMVCLNLYKLD-----NDKIKNL 875
Db 523 -----LRVEN-----SEQVLTKEEMT-----KLEDKISKVNKIKIES- 556
Qy 876 LEEAKVSTSVKTLSSSSMQPQLSLPPQDKPEVSANDDTSHSTNLSNLSKLFNLSLGN 935
Db 557 LETENNAHWLREKLKSGEACPVGSGVHHIKEGFKEVDLKALETLSKSELEGFKEK- 611
Qy 936 KNIYQELIGOKSENSEFYKILKDSDTFFNSETNFVSKKADDINSINDESKKK-LEE 992
Db 612 KPENEEIVCEASIKVEEKNIKK-----LNESINNL-----GEEFKEVSDSEKKFNLYKE 663
Qy 993 DINKLK-KTLQLSFDL--YNYKKLKLRLFDKKTKTVGYKMQIKLTLKEQLESKLSNL 1049
Db 664 KVNKENLEKIQLDDNIKDLSEBSNKIEVEYQKEIV--EKQCEKTRIVDLKSELEAIKEF 721
Qy 1050 NNPKHVLONFSVFFNKKKEAEIAETENTLTKILLKHYKGLVYKYYNGSSPLKTLSEES 1109
Db 722 N-----EVAYTIENLKAEK-----IQDFKFEKBEI----- 747
Qy 1110 IOTEDNYASLENFKVLSKLEGLKDLNKL-----EKKKLSYSSGLHHLIAELKEVIKN 1163
Db 748 -----LEKRVVRAEAGEIKDLRLNLLIRHTKEQLMDKCKRKEELSKNAELKE 798
Qy 1164 KNYTGNPSNNTDVNNALESYKFLPECTDVATVVSSESGDTLEQSOPKKPASTHVGA 1223
Db 799 KDKI-----INEXIELIK-----NKVGVL 817
Qy 1224 SNTITTSQNDVDDVIIVPIFGESEEDYDLQGVVTGEAVTPSVIDNLSKIENEYEV 1283
Db 818 DNLYEKKEIEGTIKKI-----EEQY-----NLCDKMMNEIEDKY-- 852
Qy 1284 LYLKPLAGYRSLKOLENNVMTNPNVKDIILNSRFNKNFNKVDILPYKDLTSSN 1343
Db 853 -----RKCSDEIIKYHNSLSLKDRKVNIDKLNKILMEE----- 887
Qy 1344 YVVKDPYKFLAKERDKFLSSYNYIKDSITDITIN-FANDVLGYKILSEKYSKLDLSI-- 1400
Db 888 -----KFENIEK-----AKENYLNDK--EINLLKSDV-----EKYKNELSKVNG 924
Qy 1401 -----KVIYNDKQGENEKYLPFLNN-TETLYKTVYNDKIDLVIIHLEAKVLNYYTEKSNV 1453
Db 925 AVEVLSKKLKNRKLTE-EKWIEIQNNRVEKASKAK-----ALQERSIKLEEVKNI 974
Qy 1454 EVKIKEL-NYLKTIQ-----DKLADFKNNNFVGIADL 1485
Db 975 EIKLKEGLKLLTKQELHKLKSLDLDLEKLFKGGKFEVFEVALNQL 1019
RESULT 34
MVH7_RAT STANDARD; PRT: 1935 AA.
ID MYH7_RAT
AC P02564;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
GN MYH7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=90016823; PubMed=2798112;

RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 RT "Complete nucleotide sequence of full length cDNA for rat beta
 RL cardiac myosin heavy chain";
 RN Nucleic Acids Res. 17:7529-7530(1989).
 RP [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=90133919; PubMed=2614840;
 RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
 RT Comparisons suggest a molecular basis for functional differences";
 RL J. Mol. Biol. 210:665-671(1989).
 RN [3]
 RP SEQUENCE OF 1524-1935 FROM N.A.
 RX MEDLINE=82220036; PubMed=7045682;
 RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
 RT "Molecular characterization of two myosin heavy chain genes expressed
 RT in the adult heart";
 RL Nature 297:659-664(1982).
 RN [4]
 RP SEQUENCE OF 1871-1935 FROM N.A.
 RC STRAIN=Wistar; TISSUE=Heart;
 RX MEDLINE=85179510; PubMed=6241892;
 RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
 RT "Cardiac myosin heavy chain isozymic transitions during development
 RT and under pathological conditions are regulated at the level of mRNA
 RT availability";
 RL Eur. Heart J. 5:181-191(1984).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC EMBL: X15939; AAA34065.1; -;
 CC EMBL: J00752; AAA41654.1; -;
 CC EMBL: M32698; AAA41659.1; -;
 CC PIR: S06006; S06006.
 CC PIR: A02989; A02989.
 CC HSP: P08799; LMD.
 CC InterPro: IPR000048; IQ_region.
 CC InterPro: IPR004009; Myosin_N.
 CC InterPro: IPR002928; Myosin_tail.
 CC InterPro: IPR001609; myosin_head.
 CC Pfam: PF00063; myosin_head; 1.
 CC Pfam: PF00612; IQ; 2.
 CC Pfam: PF01576; Myosin_tail; 1.
 CC Pfam: PF02736; Myosin_N; 1.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC ProDom: PD000355; myosin_head; 1.
 CC SMART: SM00015; IQ; 1.
 CC SMART: SM00242; MYSC; 1.
 CC PROSITE: PS50096; IQ; 1.

KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding.
 FT DOMAIN 1 780 MYOSIN HEAD-LIKE.
 FT DOMAIN 781 810 IQ.
 FT DOMAIN 840 1935 COILED COIL (POTENTIAL).
 FT NP_BIND 178 185 ATP.
 FT DOMAIN 655 677 ACTIN-BINDING.
 FT DOMAIN 757 771 ACTIN-BINDING.
 FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 695 695 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 705 705 ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 1529 1531 IRK -> VRR (IN REF. 3).
 FT CONFLICT 1731 1731 D -> H (IN REF. 3).
 FT CONFLICT 1784 1784 N -> K (IN REF. 3).
 FT CONFLICT 1851 1851 T -> N (IN REF. 3).
 FT CONFLICT 1858 1858 R -> K (IN REF. 3).
 FT SEQUENCE 1935 AA: 223082 MW: C8376C324A7BD82B CRC64;
SQ
 Query Match 4.2%; Score 351.5; DB 1; Length 1935;
 Best Local Similarity 18.9%; Pred. No. 0.00011;
 Matches 286; Conservative 265; Mismatches 589; Indels 377; Gaps 62;
 QY 28 LVYKLEALEDAVLGYSLFQEKVMVNEGTSCTAVTSTPGSKGVSASGGSGS 87
 DB 594 LOKNKDPLNETVV--GLYQKSSLLKLLSLNFANYAGADAPVDKGVK-----GK 637
 QY 88 VASGGSVASGGSVASGGS-----NSRRTNPSDNSSDAKSYADLKHRYNVLITIKELK 143
 DB 638 AKKGSFQVTSALAHRENKLNKMTNLRSTHP-----HFVR--CIIPNETK 679
 QY 144 YPOLFD--LTNHLTLCDNIHGFKYILGDYEEINELLYK-----LN-----FYFD 186
 DB 680 SPGVMDNPLVMHQLRCNGVLEIGIRICRKGFP--NRILYGFQRYRILNPAAPPEQGFID 737
 QY 187 LLRA--KLMDVCANDYCYQIPFN-----KIRANELDVLK 218
 DB 738 SRKGAELKLSGLSDIDHNYKFQKGTQVFFKAGLLGLEEMRDERLSRIITRQAQSGVLS 797
 QY 219 KLVEFGYKPLD-----NIKDNVG-----KMEDYIK--KNKKTIEINEL 255
 DB 798 RHEF--KLLERDLSLIITQWIRAPMGVKNWPMWKLKPKPLKASATEKEMANKEE 855
 QY 256 IEESKTTIDKNKNATKEEKK--KLYQAQYDLSYKQLEEAHNLISVLEKRTIDTLKKNE 313
 DB 856 FGRVKDALEKSEARKELEKVMYSLLOEKNDLQL---QVQAEODNLADAEERCDQLIKNK 912
 QY 314 -----NIKELDKINEIKNPPANSNGTNTWLLDKNKKIEEHEKEIKEAKTIKFNIDSL 368
 DB 913 IQLEAKVKEMTERLEDEE-----EMNAELTAKKRKLEDECSSELKRDID-- 955
 QY 369 FTDPLELEYLREKNKNIDISAKVE--TKESTEPNPNYPTVPLSYNDINNALNELNSFG 427
 DB 956 -----DLEITL-----AKVEKEKATE-----NKVKNLTEEMAGLD 986
 QY 428 DLINPFDYTKPSKNYITDNERKKFINETKEIKIKIEKKIESDKKSYEDRSKSLNDITRE 487
 DB 987 EII--VKLTKE-----KKALQEAHQ--ALDDLOAEEDK-----VNTLTKA 1023
 QY 488 YEKLLNEIYDSKFNENNIDLTNFEKMMGKYSYKVE---KLTHHTNTFASYSKHNLEKLT 544
 DB 1024 KYKLEQQVDD--LEGLSDQDKKVRMDLERAKRKLEGLDKLT--OESTIMDLENKQQLDERL 1080
 QY 545 KALKY-----MEDYSLRNTIWEKELKYKNLSIKENIETLTVENTIKKDEQLFEKK 596
 DB 1081 KKKDFELNALNARIEDEQALGSQLKKLQELQARIELEEELE--AERTARAKVEKLRS 1138
 QY 597 ITKDNENKPEKILEVSDIVKQVQKVLNNK--IDELKKTQLTLKVKELKHNTHVPSNYK 654
 DB 1139 LSRELEIEISERLEEAAGGATSVQIE-----MNKREAEFOKMRDRDLLEATLQHEA---TAA 1190
 QY 655 QENKQEPYLVILVKKKIDKLKVFMPKVESLINEEKKNIKTEGOSNSEPSTEGEITGOAT 714

Db 1191 ALRKKHADSVAELEQIDNLO-----RVQKLEKESEFKLE--LDDVTNNMEQIIIAKAN 1244
QY .715 ----TKPGQAGSALGSDSVQAOAQOKOAOQPPVPVPEAKAQAQVTPPPAPVNNKTN 768
Db 1245 LEKMCRTLEDQWNEHRSKAEETQSRVNDLTQRAKLQTEGELSQOLDEKEALISQLTR- 1303
QY 769 VSKLDYLEKLEFLNTSYCHIKYILVSHSTNMEKILQYKITKEESKLSODPDLDFN 828
Db 1304 -GKLTYTQOEDL-----KQLEEEVKAKNALAHALQSAHDCDLDLREQVE 1348
QY 829 IQNNIPVMYSPDLSNLSQLFMELIYKEMVCNLYKKD-NDKIKNLEEAKKYSTSVK 887
Db 1349 ETEKAELQRLVSKANSEVAQWRKYETAQRTTELEEAQKLAQRLQDAEAEAVN 1408
QY 888 TLSSSSMPLSLTPQDKPE---VSANDDTSHSTNLSNLSKLFENILSGKKNKNIQEELIG 944
Db 1409 A-KCSSLEKTKHRLQNEIEDLMVDYERSNAAAALDKKORNFDKILVEWKQK--YEEQS 1465
QY 945 QKSS-----ENYEKILKDSDFYNESFTNFVKSADDTINSLND----- 983
Db 1466 ELESSQKEARSLSLTFELKKNAYEESLEHLETFKREN-----KNLQEEISDLTEQLGSTG 1520
QY 984 -----ESKRKKLEEDINKLKT-----QLSPDLNKKYKIKLER-LF 1019
Db 1521 KSIHELEKIRKOLEAEKLEQSALEAEASLEHBEKGILRAQLEF---NQKAEIERKLA 1577
QY 1020 DKKKTGVGYKM-QIKKLTLLKQESKLSLNNPKHVLQNFVFFNKKKKEAIEAETNTL 1078
Db 1578 EKDEMEQAKRNLRLVDSLOTSLDAETRSNEALRV-----KKMEGDLNEMETQL 1629
QY 1079 ENTKILLKHGKLVKYNGESSPLKLTSEESIQTEDNYASLENFVKLSKLEGLKDNLNL 1138
Db 1630 SHANMAEAQKQV-----SLQSLKDTQIQDDAVRANDD-----LKENIAI 1673
QY 1139 EKKKLSYLSGLHLHILAEKLEKIKKNYTGNSPSNNTDV-----NNALESYKFL-PE 1191
Db 1674 VERRNNLQAELEELRAVVEQTERSKLAEOELIETSERVOLLHSONTSLNOKKMDAD 1733
QY 1192 GTDAVTVYSESGDFLEQSQPKPASTHVGAESNTITTSQ-----NVDDEVDDV 1240
Db 1734 LSOLTEVEAEVQECRAEAEAKKAITDAAMMAELKKEQDTSALHMKNNWEQTIKDL 1793
QY 1241 IIVPIGSEEDYDQGV-VTGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLKQ 1299
Db 1794 -----QHRLEAEQIALKGKQKQLEARVRELENELE-----AEQKR 1832
QY 1300 LENNVMTNVNVDILNRFKNRENFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKD 1359
Db 1833 NAEVKGMRKSERRIKELTYQTEEDRKNLLR-----LQDLVDKQLQKVKAYKQAEAE 1887
QY 1360 KF---LSSYNYIKDSID 1373
Db 1888 QANTNLSKFRKQVHOLD 1904
RESULT 35
MYH7_HUMAN STANDARD; PRT; 1935 AA.
AC P12883; Q14904; Q16579;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MYHC-beta).
GN MYH7 OR MYHCB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91065634; PubMed=2249844;
RA Jaenike T., Diederich K.W., Haas W., Schleich J., Lichter P.,
RA Pfordt M., Bach A., Vosberg H.P.;

RT "The complete sequence of the human beta-myosin heavy chain gene and
a comparative analysis of its product.";
RL Genomics 8:194-206(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90301496; PubMed=2362820;
RA Liew C.-C., Sole M.J., Yamauchi-Takahara K., Kellam B.,
RA Anderson D.H., Lin L., Liew J.;
RT "Complete sequence and organization of the human cardiac beta-myosin
heavy chain gene.";
RL Nucleic Acids Res. 18:3647-3651(1990).
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN [5]
RP SEQUENCE OF 1310-1935 FROM N.A.
RX MEDLINE=86176778; PubMed=2421254;
RA Saez L., Leinwand L.A.;
RT "Characterization of diverse forms of myosin heavy chain expressed in
adult human skeletal muscle.";
RL Nucleic Acids Res. 14:2951-2969(1986).
RN [6]
RP REVISIONS.
RA Leinwand L.A.;
RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1410-1935 FROM N.A.
RX MEDLINE=88299163; PubMed=2969919;
RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
RT "Molecular cloning and characterization of human cardiac alpha- and
beta-form myosin heavy chain complementary DNA clones. Regulation of
expression during development and pressure overload in human
atrium.";
RL J. Clin. Invest. 82:524-531(1988).
RN [8]
RP SEQUENCE OF 785-1935 FROM N.A.
RTISSUE=Skeletal muscle;
RX MEDLINE=90235862; PubMed=1691980;
RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
myosin heavy chains.";
RL Eur. J. Biochem. 189:55-65(1990).
RN [9]
RP SEQUENCE OF 1393-1935 FROM N.A.
RX MEDLINE=87192738; PubMed=3032769;
RA Jandreski M.A., Liew C.-C.;
RT "Construction of a human ventricular cDNA library and
characterization of a beta myosin heavy chain cDNA clone.";
RL Hum. Genet. 76:47-53(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=96039076; PubMed=8533830;
RA Arai S., Matsuoka R., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
RA Kimura M., Imamura S.-I., Furutani Y., Joh-O K., Kawana M., Takao A.,
RA Hosoda S., Momma K.;
RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
hypertrophic cardiomyopathy.";
RL Am. J. Med. Genet. 58:267-276(1995).
RN [11]
RP VARIANTS CMH1 GLU-256 AND ARG-741.
RX MEDLINE=93248216; PubMed=8483915;
RA Panapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;
RT "Missense mutations in the beta-myosin heavy-chain gene cause central
core disease in hypertrophic cardiomyopathy.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).

[illegible]

Db	1110	NLVARETIDTQKDDLCRFADILG-AEROCERLYSWASGKR-----HQLETEATM	11150
Qy	1284	LYLKPLAGVYRSLKQLQENNVMVTFNVNVDILNSRFNKRENFKNVLESOLIPYKDLTSSN	13443
Db	1159	CDAKTLVRNNEVEKMMMSRIGEVQ-QLRDFL-----DNLKASKTSDTQTTEL----	1206
Qy	1344	YVVKDPKYFLNKEKDKFLSSYNYIKDSI---DTINFANDVLGYKKILSKYKSDLDLSI	14000
Db	1207	---ENKFEELDVE-----WNHLEELRRRREVOLN-----DSM	1235
Qy	1401	KKYINDKQGENEYKLPFLNNIETLYKTVNDKIDLFVHLEAKVLNYYEKSNNVEVKIKEL	1460
Db	1236	SOILIDSO-----FDFIRMIKERTEA-----LEADV-----EVSN-----	1266
Qy	1461	NLYKTIOQKLADFKNNNFVGIADLSTFYN-HNNLLTKFLSTGMVFENLAKTV---LSNL	1516
Db	1267	---KTKPDDIERQKRHD-----ELASDLNDHYTYEDEFLLNGQVEEEKVQVIRHLNSSL	1318
Qy	1517	LDGNLOGLMNLISHCQVKYKQCPQNSGCFRHLDD	1548
Db	1319	IESSTTRQKSLAQ-----EC-QKSLRFELLD	1343
RESULT 37			
MYH6_MESAU			
ID	MYH6_MESAU	STANDARD;	PRT; 1939 AA.
AC	P13539;	O60562;	
DT	01-JAN-1990	(Rel. 13, Created)	
DT	15-DEC-1998	(Rel. 37, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last annotation update)	
DE	Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).		
GN	MYH6.		
OS	Mesocricetus auratus (Golden hamster).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;		
OC	Mesocricetus.		
OX	NCBI_TaxID=10036;		
RP	SEQUENCE FROM N.A.		
RP	SPRAIN-FIB; TISSUE=Liver;		
RC	MEDLINE=5115033; PubMed=7815459;		
RA	Wang R., Solt M.J., Cukerman E., Liew C.-C.;		
RT	"Characterization and nucleotide sequence of the cardiac alpha-myosin		
RT	heavy chain gene from Syrian hamster."		
RL	J. Mol. Cell. Cardiol. 26:1155-1165(1994).		
RP	[2]		
RP	SEQUENCE OF 1630-1939 FROM N.A.		
RX	MEDLINE=86205859; PubMed=3458174;		
RA	Liew C.-C., Jandreski M.A.;		
RT	"Construction and characterization of the alpha form of a cardiac		
RT	myosin heavy chain cDNA clone and its developmental expression in the		
RT	Syrian hamster";		
RL	Proc. Natl. Acad. Sci. U.S.A. 83:3175-3179(1986).		
CC	-!- FUNCTION: MUSCLE CONTRACTION.		
CC	-!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2		
CC	HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)		
CC	AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).		
CC	-!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.		
CC	-!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING		
CC	CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,		
CC	CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.		
CC	-!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY		
CC	ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY (BY		
CC	SIMILARITY).		
CC	-!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT		
CC	MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE		
CC	SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED		
CC	SUBFRAGMENT (S2).		
CC	-!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE		
CC	MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.		
CC	-!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.		
CC	-!- SIMILARITY: CONTAINS 1 IQ DOMAIN.		

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; L15351; AAB59701.1; -;
 DR EMBL; M12995; AAA37081.1; -;
 DR HSP; P08799; IAMD.
 DR InterPro; IPR000048; IO_region.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding.
 FT DOMAIN 1 782
 FT MYOSIN HEAD-LIKE.
 FT DOMAIN 783 812
 FT IQ.
 FT DOMAIN 842 1939
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1439 1443
 FT POLY-ALA.
 FT NP_BIND 178 185
 FT ATP (POTENTIAL).
 FT DOMAIN 657 679
 FT ACTIN-BINDING.
 FT DOMAIN 759 773
 FT ACTIN-BINDING.
 FT MOD_RES 129 129
 FT METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 597 597
 FT ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 707 707
 FT ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 1633 1633
 FT Q -> L (IN REF. 2).
 FT CONFLICT 1651 1651
 FT H -> Q (IN REF. 2).
 FT CONFLICT 1686 1687
 FT EL -> DV (IN REF. 2).
 FT CONFLICT 1693 1693
 FT V -> G (IN REF. 2).
 FT CONFLICT 1844 1844
 FT K -> R (IN REF. 2).
 FT CONFLICT 1879 1879
 FT A -> T (IN REF. 2).
 FT CONFLICT 1885 1885
 FT E -> Q (IN REF. 2).
 FT CONFLICT 1907 1907
 FT E -> V (IN REF. 2).
 FT CONFLICT 1928 1928
 FT D -> N (IN REF. 2).
 FT CONFLICT 1933 1935
 FT OKM -> KR (IN REF. 2).
 SQ SEQUENCE 1939 AA; 223626 MW; DBC8297DFE83115A CRC64;

Query Match 4.1%; Score 349; DB 1; Length 1939;
 Best Local Similarity 19.4%; Pred. No. 0.00014;
 Matches 297; Conservative 255; Mismatches 577; Indels 400; Gaps 65;

Qy 28 LVKLEALEDAVLTCYLFQREKXVNLNEGTSCTAVTTTPSGKGVASGGSGVASGGS 87
 Db 595 LEKNKDLNETVW-----GLYQSSSLK-----MATLFT-----YASADAGDSGKRG 639
 Qy 88 VASGGSVASGGSGVASGGSG-----NSRRTPNSDSSDSDAKSVADLKHVRVNLTIKELK 143
 Db 640 KKKGSFQTVSAHLRENLNKMLTNLRTTHP-----HFVR--CIIPNERK 681
 Qy 144 YPOLFD--LTNHLTCLDNIHGFYKLYIDGYEINELLYK-----LN-----FYFD 186
 Db 682 APGVMDNPLVMHQLRCNGVLEGIRCKRGGP--NRILYGDPRQRYRLNPAAPBEGQFID 739
 Qy 187 LLRA--KLNDVCANDYCOIPN-----LKRANELDVLK 218
 Db 740 SRKGAELSLSDIDHNGYKFGHTKVFVKAGLLGLLEEMRDERLSRIITRIQAARGQLM 799
 Qy 219 KLVFYGRKPLD-----NIKDNVG-----KMEDYK--KNKKTININEL 255
 Db 800 RIEP--KKNVRRDALLVIQNIIRAFMGVKNWPMKLYFKIRPLKLSAETKEKMANMKEE 857

Qy 256 IEESKTKDKNKNATKEEEK--KLYOQYDLSIYNKOLEEAHNLISVLEKRTDKKNE 313
 Db 858 FORVSELEKSEKARKELEKMYSLLOEKNDLQF---QVQAEQDNLNDAEERCDQLIKNK 914
 Qy 314 -----NIKELDKLINEIKNPPANSNTPTLLDKNKKEHEKEIKEIAKTKFNIDSL 368
 Db 915 IQLEAKVEMTERLEDEEE--MNA-----ELTSKKRKLED--ECSELKDOI----- 956
 Qy 369 FTDPLELEYLREKKNKIDISAKVE--TKESTEPNEYPNGVTYPLSYNDINNALNELNSFG 427
 Db 957 --DDELELT-----AKVEKEHATE-----NKVKNTEEMAGLD 988
 Qy 428 DLINPFDYTKPSKNYI-----TDNERKFFINEIKEIKIEKKKIESOKKSYEDRSKSL 481
 Db 989 ELIAKLTREKKAQEAHQALDQAEDEKVNLTLSKVKLE--QQVDDLEGSLQEKKYR 1047
 Qy 482 NDIKEYEKL-----LNEIYDSKFNNDIDITNFKMMGKRYSYKVEKLTHTHINTFASVENS 538
 Db 1048 MDLERAKRKLEGLDNTVQESIMDLENDKLQLE-----EKLKKKEFDISOQNSK- 1095
 Qy 539 NLEKLTALKYMEDYSLRNIVVEKELKYYKNLISKIENEIETLVENIKKDEEQLFEKKIT 598
 Db 1096 -----JEDEQALALQLOKKLKENQARIEEELEEE--AERTARAKVEKURSDIT 1142
 Qy 599 KDNKPDEKILEVSDIVKQVQKVLMMNK--IDELKKTOLILKNVLELKHNIHVNPNSYKOE 656
 Db 1143 RELEISERLEAGGATSQIE---MNMKREAEFQKMRDLEEAATLQHEA---TAAAL 1194
 Qy 657 NKQEPYLLIVLKKETDKLVFMPKVESLINEEKKNIKTGQSDNSEPSTEGEITQATTK 716
 Db 1195 RKKHADSVAAELGEQIDNLQ---RVKQKLEKSEKSEFKLE--LDDVTSNMEEQIKAKANL- 1247
 Qy 717 PQQAAGSALGDSVQAQAQEQKQAPVPVPVPEAKAQPPTPPAPVNNKNVNSKLDYLE 776
 Db 1248 --EKVSRILEDOANERYVKLEESQRSLSNDFTTQRAKLO-----TENGEARQLE 1294
 Qy 777 KLYEFLNTSYCHKYILVSHSTMNE-----KILKQYKITEESKISS-----C 820
 Db 1295 E-----KEALISQLTRGKLSVTQOMEDLKROLEEGKAKNALAHALQASRHC 1342
 Qy 821 DPLDLLFNIQNNIPVYSMFDSLNLSOLFMEIYEKEMVCNLYKLD--NDKIKNLLSEA 879
 Db 1343 DLLREQYEEEMBAELQRLVLSKANSEVAQWRTYETDAIORTTEELEAKKLAORLODA 1402
 Qy 880 KKVSTSVKTLSSSSMOPSLTPQDKPE---VSANDDTSHSTNLNLSKLFENILSGNKK 936
 Db 1403 EEAVEAVNA--KCSSLEKTKHRLQNEIEDLWVDVRSNAAALDKKORNFDKILAEWKOK 1461
 Qy 937 NIYQELIGQKSS-----ENFYKILKSDTFYNESFTNFVKSKADDDINSL 981
 Db 1462 --YEESQSELESSQKARSLSLTFELKKNAYEESLEHLETFKREN-----KNLQBEISDL 1514
 Qy 982 ND-----ESKRRKLEEDINKLKKTL-----QLSFDLYNKVK 1012
 Db 1515 TEQLGEGGKNVHEKVRKQLEVEKMELOSLAEAEASLEHEEGKILRAQLEF---NQIK 1571
 Qy 1013 LKLER--LFDKKTGCKYKM--QIKKLTLLKLEOESLNSLNNPKHVLQNFVFEFNKKEAE 1070
 Db 1572 AEIERKLAEKDEEMQAKRNHLRVVDSLSQTSLEAETRSNEALRV-----KKKMEGD 1623
 Qy 1071 IAEETNTLENTKILL-----KHYKGLVYNGYNGESSPLKTLSEESIQETEDNYASLENPKV 1126
 Db 1624 LNMEIQLSQANRIASEAQKHLK-----NAQAH-----LKDTQLQDLDALHAND- 1668
 Qy 1127 KLEGLKDNLNLEKKKLSYLSGLHLLIAELKEVKNKNYTCNSPSENNTDV-----N 1179
 Db 1669 -----LKNIAIVERNTLLQAELEELRAVVEQTRSKLABQELIETSERVOLLHSQNT 1723
 Qy 1180 NALESYKGFPLPGTGVATVWSESGSDTLEQSQPKKPPASTHVGAESNTITTS- 1231
 Db 1724 SLINQKKMEADLTQLOTEVEEAECRNAEAKAKAITDAAMMAEELKKEODTSAHLER 1783
 Qy 1232 ---NVDEVEDDVIIVPFCSEEDYDLDQGV--VTGEAVTPSVIDNLSKIENEYVLYLK 1287

```
Db 1784 MKNMEQTIKDL-----QHRLEAEQIALKGCKKQLOKLEARVRELENELE-----1829
QY 1288 PLAGVYRSKLEKOLENNVMTFNVNVDILNSRKNRKNFNKLVESDLIPYKDLTSSNVVVK 1347
Db 1830 -----AQKRNAESYKMKRSERIKELTQTEEDKRN-----LVRLODLVDKQLQKV 1877
QY 1348 DPYKFLNKKRDKF-----LSSYNYIKDSID 1373
Db 1878 KAYKQAEAEAEQANTLSKFRKQVQHELD 1906

RESULT 38
MYH3_CHICK
AC P02565; STANDARD; PRT; 1940 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, fast skeletal muscle, embryonic.
GN MYH3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87194881; PubMed=3571266;
RA Molina M.I., Kropp K.E., Gulick J., Robbins J.;
RT "The sequence of an embryonic myosin heavy chain gene and isolation
of its corresponding cDNA.";
RL J. Biol. Chem. 262:6478-6488(1987).
RN [2]
RP SEQUENCE OF 1502-1940 FROM N.A.
RX MEDLINE=83161144; PubMed=6833296;
RA Jakovitsky C.J., Umeda P.K., Sinha A.M., Elzinga M., Tong S.W., Zak R.,
RA Jakovits S., Rabinowitz M.;
RT "Cloned mRNA sequences for two types of embryonic myosin heavy chains
from chick skeletal muscle. I. DNA and derived amino acid sequence of
light meromyosin.";
RL J. Biol. Chem. 258:5196-5205(1983).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2)
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V00430; CA23712.1; .
DR EMBL; J02714; AAA48972.1; .
DR PIR; A02990; A02990.
DR PIR; A29320; A29320.
DR HSSP; P13538; 2MYS.
DR InterPro; IPR000048; IQ_region.
```

```
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
FT DOMAIN 786 815 IQ.
FT DOMAIN 844 1940 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP.
FT DOMAIN 660 682 ACTIN-BINDING.
FT DOMAIN 762 776 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 700 700 ALKYLATION (SH-1).
FT MOD_RES 710 710 ALKYLATION (SH-2).
FT VARIANT 379 379 G -> D.
FT CONFLICT 1547 1547 T -> A (IN REF. 2).
FT CONFLICT 1913 1915 ERA -> GRT (IN REF. 2).
SQ SEQUENCE 1940 AA; 222816 MW; C34833D75B04DF2 CRC64;

Query Match 4.1%; Score 348; DB 1; Length 1940;
Best Local Similarity 19.1%; Pred. No. 0.00015;
Matches 299; Conservative 262; Mismatches 520; Indels 482; Gaps 71;

QY 37 DAVLTGYSLFQKEKVMVNEGTSCTAVTTSTPGSKGVSASGSGSVASGSGSVAS 96
DB 590 DNYITGW--LEKNKDPLNETVGLYQKSLKTLALFAS--VGGAAESGAGKGGKKK 645
QY 97 GGSVASGGS-----GNSRRTPNSDSDAKSYADLKHVRVNYLLTIKELKYPQ 146
DB 646 GGSFQTSALFRENKLMNSLRSTHP-----HFVR--CLIPNETKTPG 687
QY 147 LF-DLTNHLTLCDNIHGFYKIDG-----YEINELLYKLNFFVDLLRA- 190
DB 688 AMEHELVLHQLRCNGVLEIGRICRKGFPRIILYADFQRYKVLNLSAIPGQFIDSKAS 747
QY 191 -KLVDCANDYCOIPE--NLKIRANELOVLK-----K 219
DB 748 EKLGSIDVDHTQYKPGHTKVFVKAGLLGLEEMRDEKLAQLITRQACRGFLMRVEFK 807
QY 220 LVFGYRKPLDNIKDNVG-----KMEDIYK--KNKKTIEINELIEESKKT 262
DB 808 KMMERRESIFCIQYNVRAFNMVKNVPMKLFKKIPLLKSAESEKEMANNKEFEKTKEE 867
QY 263 IDKNKNATVEEKK--KLYQAQYDLSIYNKOLEEAINLSVLEKRTDITLKKNE-----NI 315
DB 868 LAKEARKLELEKVMVSLQEKNDLQ--QVQAADGLADAEERCDQLIKRIQLEAKI 924
QY 316 KELLDKINIKNPPPPANSNGTPTNLLDKNKKTEEHEKEIKETAKTIKFNIDSFTDPLEL 375
DB 925 KELTERAEDEE--MNA-----ELTAKRKKLED---ECSELKKDI-----DDEL 964
QY 376 EYLLREKKNKIDISAKVE--TKESTEPNEYPNGVTPYPLSYNDINNALNELNSFGDLINPFD 434
DB 965 TL-----AKVEKEKHATE-----NKVKNLTEEMAALDETIACL- 997
QY 435 YTKESKNITYONERKKNFNETKEIKIEKKKIESDKSYEDRSKSLNDITREYKLLNE 494
DB 998 -TKE-----KKALQEAHQ--TLDDLOAEEDK-----VNTITRAKTKLEQ 1035
QY 495 IYDSKFNENNIDLTPNPEKMMGKRYKYKVE--KLTHNTFASVENSKHNLKLTALKY-- 549
DB 1036 VDD--LEGSLEQEKKLURMDLERAKRKLKGLDLMQESTM--DLENDKQQLDEKLUKDKDFEI 1092
```

```
Qy 550 -----MEDYSURNIVWEKELKYKNLSKISKEIENETIVENIKKDEQLFEKKITDKENK 603
Db 1093 SOIQSKIEDQALQMOLQKKIKELQARIEEETEE--AERTSRAKAHKRADLSRELEE 1150
Qy 604 PDEKILEYSDIVKQOVQVLLMNK--IDELKKTOLILKNVELKH----- 645
Db 1151 ISELEEAGGATAAQID-----MNRKREAEFOKMRDLBEATLQHEATAAALRKKHADSTA 1206
Qy 646 -----NIHPNSYKQENKOPPYLIVLKKEID-----KLKVFMPKVESLINEE 688
Db 1207 DVGEQIDNLQRYQKLEREKSE-----LKWEIDDLASNMESVSKAKANLEKMCRSLEDQ 1260
Qy 689 KNIKTEGQSDNSESTGEITGQAT-----TPGQOAGSALEDSDVQAQEQKQAPVPV 745
Db 1261 LSEIKTK---EEOQRTINDISAQARLQTESGEYSRQVEEKDALISQSRGKA----- 1312
Qy 746 VPVPEAKAQVPTTPAPVNNKTNVSKLDYLEKLYEFLNTSYCHKYILVSHSTNMEKILK 805
Db 1313 -----FTQOIEEL----- 1320
Qy 806 QYKITKEESKLSSCDPLDLFNIGNONIPVMYSMD-----SLNNSLSQLFMFI----- 854
Db 1321 --KRHEEBEIRAKKC-PAHALQSARHCDLLREQVEEBOEAKGELQALRSKRANSEVAQWR 1377
Qy 855 --YKEMVCNLYKLD-NDKTNLLEBAKKVSTSVKTLSSSSMQPLSLTPQDKPE---VS 908
Db 1378 TKYETDAIQTTELEBEAKKKAQRLQDAEHEAVNS-KCASLEKTKORLQNEVEDLMID 1436
Qy 909 ANDDTSHSTNLNLSLKLFNILSLGKNKIYOEL-----IGOKSS-----ENFEY 953
Db 1437 VERSNAACAALDKKQKFDKILSEWKQ--YEETOAELEASQKESRSSTELFKMKNAYE 1494
Qy 954 KILKSDDTFYNESTFNFYKSKADDINSND-----ESKRKKLEEDINKL--- 998
Db 1495 ESLDLHLETKREN-----KNLQEISDLTEQIAEGGKAIHELEKVKQIEQEKSELOPAL 1549
Qy 999 -----KTLQSLFDLYNKKYKLLER-LFDKKTIVGKYK-QIKKTLLEKEQLES 1044
Db 1550 EEAASLEHEGKILRVQLEL-NOVKSIDRKAIEKDEIDQLKRNHLRVVDSMQSTLDA 1608
Qy 1045 KINSLNPKHVLQNSVFNFKKKEAEI-----AETENTLENTKILLKHYKG 1090
Db 1609 EIRS-----RNEALRKKMEGDINEIETQLSHANRQAAEQKNLRTQGVLK--- 1656
Qy 1091 LVKYNGESSPLKLTSESIOTEDNYASLENFKVLSKLEGKLDKNLNEKKKLSYLSGL 1150
Db 1657 -----DTQIHLDDALRSQED-----LKEQVAMVERRANLLQAEI 1690
Qy 1151 HHLTAELKEVIKNKNYTGNSPSENNTDV-----NNALESYKKLPFGTDTVATVVSSEGS 1204
Db 1691 EELRAALEQTERSKRVASQELLDASERVOLLHTQNTSLINTKKLE--SDISQIQSEM-E 1747
Qy 1205 DTLEQS-----OPKKPASTHVGAESNTITSO-----NVDDVDDVLIIVPFGES 1249
Db 1748 DTIQEARNAEKAKAITDAAMMAEELKKEQDTSAHLERMKKNLDQTVKDL----- 1798
Qy 1250 EEDYDDLQGV-VTGEAVTPSVIDNLSKIENEYEVLYLKPLAGVYRSLSKOLE--NNVMT 1306
Db 1799 QHRLEDEAQLGALKGKKQIQIKLEARVREGEVDA-----EQKRSAAEVKGVK 1847
Qy 1307 FNVNVKDLNRFNKNRKNFKVNLWSDLPYKDLTSSNVVVKDPYKFNKERRKDFLSSYN 1366
Db 1848 YERRYKEL---TYOSEEDRKNVLR-----LQDLVDKLMQVKYSKQRAEEAE--LSNVN 1897
Qy 1367 YIK 1369
Db 1898 LSK 1900
```

RESULT 39

SMCL_YEAST

ID SMCL_YEAST STANDARD; PRT; 1225 AA.

AC P32908;

```
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosome segregation protein SMC1 (DA-box protein SMC1).
GN SMC1 OR CHL10 OR YFL008W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103320; PubMed=8276886;
RA Struninikov A.V., Larionov V.L., Koshland D.;
RT "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT protein is required for nuclear division and defines a new ubiquitous
RT protein family."
RL J. Cell Biol. 123:1635-1648(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae."
RL Nat. Genet. 10:261-268(1995).
CC -!- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00602; AAA16595.1; -;
CC EMBL; D50617; BAA09230.1; -;
CC PIR; A49464; A49464.
CC PIR; S41804; S41804.
CC SGD; S0001886; SMC1.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC Mitosis; ATP-binding; Coiled coil; Nuclear protein.
CC NP_BIND 33 40
CC DOMAIN 173 489
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC ALA/ASP-RICH (DA-BOX).
CC S->L: IN TS MUTANT SMC1-2.
CC MUTAGEN 173 173
CC MUTAGEN 458 458
CC N->D: IN TS MUTANT SMC1-1.
CC SEQUENCE 1225 AA; 141279 MW; B504017AA0ECCAC8C CRC64;
SQ
Query Match 4.1%; Score 347; DB 1; Length 1225;
Best Local Similarity 20.3%; Pred No. 9.6e-05;
Matches 284; Conservative 248; Mismatches 505; Indels 364; Gaps 69;
Qy 203 IPFNLKIRANEL--DVLKKLVFGYKRPDLNLIK-----DNVGMED-----YIK-----KNKK 247
Db 46 ISFVLGVRSNHLRSNLIKDLI--YRGLNDNSDDYDNEGAASSNPQSAAYKAFYQKGNK 103
Qy 248 TIENINELIESKTTIDKNKNATKEEBKKLYQAOYDLSIYNKO-LEBAHNLIISVLEKRI 306
Db 104 LVELM-----RIISRNGDTSYKIDKTKVSYKDYISIFLENILIKAKNPL-VFQGDV 154
```

Qy	307	DTLKKKNTEKLLDKINETKNNPPPPANNTPNTLLDKWKKIEEHKEKETEIAKTIKCNID	366
Db	155	EQIAAQPVP-ELSRMPEEV-----SGS-----IQYKKEYEELKEKTEKLSKATESIK	201
Qy	367	SLFTDPLELEYILREKNNKIDISAKVEYKESTEPNEYPNGVTYPLSYNDINNALNSF	426
Db	202	NRRIHGEUKTYKEGINNVEYRKODKK-----NELQKF	236
Qy	427	GDILNPDYTKPSKNITYD--NERKKFINEIKKIKIEKKKIBSDKKSVEDRS-----	478
Db	237	QALWQL--YHLEQOQKEELDKLSALNSELISSLKGKINNEMKSLQBSKSFYKESAVISKQ	294
Qy	479	-KSLNDITKEYEKLNEIYDSFPNNNIDJTNPEKMWGRKRYKSYKVEK-----LTHHNT	529
Db	295	KSKLDYIFDKELVSDLRILKVP-----QQAAGKRISH--TEKRTIESLQKDLQKQKT	345
Qy	530	FASYENSKHLEKTLKALYMED--YSLRNI-----VVEKELKYKKNLSK-----I	574
Db	346	IV--ERFETQLKVVYTSKAFBEEIKQSARNYDKFKLNENDLTKYCNLHEKYLTEGGSIL	403
Qy	575	ENBIETLVENIKKDEQL-----FEKTKTKDNKPDEKI--LEVSADI--VKVQVKVL	623
Db	404	EEXIAVLNNDKRIQIELELERNKRADISKRRITEELSITGEKLDLNDLRVSLNEKNAL	463
Qy	624	LMKKIDELKQTQILNLKVELKINIHVPSYK-----QENKQPPYILVLKKEID	672
Db	464	HTERLHELAKLO---SDIESANNQEVDLNFKLRETLVLIKIDOLSANORETMKERKLENIA	520
Qy	673	KLAVEMPKVESLINEEKKNIKTGOSDNSEPSTEGEITQOATTPGOQAGSAL--EGDSV	730
Db	521	MLKRFPGVGLVH-----DCHPKKE-----KYGLAVSTILGKNFDS	559
Qy	731	QAO-----AQE-----QKOAQPPVPVPEAKAQAQVTPPAP--VNNKTNVSKLDYLEKLY	779
Db	560	IVENLTVAQECIAFLKQKAGTASFIPLDTIETELPTLSLPDSQDYILSINAIDY-EPEY	618
Qy	780	EFLNTSYCHKYLIVSHSTWNEKILKQYKITHKEESKLSKSCDPLDLFNTQONNIPVMSY	839
Db	619	E-KAMOYVCGDSIIC--NTLN--IAKDLKWKKGIRGKLVITIE-----GALIHK	661
Qy	840	FDSLNNLSQLFMEIYEKKEWCVNLKLDN-----DKIKNLE-----EAKKVTSVKTL	890
Db	662	AGLWTGGISCDANNRWDXEYQSLMSLKDKLIIQIDELSGORSNSTRAREVENSVLIN	721
Qy	891	SSSQPLSTLPQDKPVSANDTSHSTNNLSKLFIENILSGKNKNIYQELIGQKSS--	948
Db	722	SDIANLRQTVOOK-----RSLDENRLBTKYHNDLIEKIQPKITEL	763
Qy	949	-----ENFYEKILKSDTFVNESFTNFKVSKADDINSLNDESKR--KLEEDINKLK	999
Db	764	KKLDLLENTKMLVKEKALQNNIFKETSIGITFIKEYENHSGELMROQSKLELOLQK	823
Qy	1000	TLQLSFDLYNKYIKLERLFDKKTGVYKMQIKKLTLLKQLBSKLSNLNPKHVLQNF	1059
Db	824	QI---LTVENKLOFETDRL--STTORRYEKAQKDL---ENAQVEMKSLBEEQYAI---	870
Qy	1060	SVFENKKKEAETAENTLNTKYLKHKYGLVYKNGESSPLKTLSPESIQTEDNYASL	1119
Db	871	-----EMKIGISIEKLEEHK--NHLDELQKFVTKQSELN--SSSEDILEDMN---	913
Qy	1120	ENFKVLSLGBGLKLDNLNLEKKKLSVLSGLHLHIAELKEVINKKNYTG--NSPSENNDV	1178
Db	914	SNLOVLKRRDGLKE--DIEKFDLBERVA-----LKNCISNINIPISSETTI	959
Qy	1179	NNALESYKKFLP-EGTDVATVYSEGSDTLEQSPKPKASTHVGAESNTITTSQNVDDV	1237
Db	960	DD-----LPISSTDNEAIIISIDINYGKLPKK-----YKENNTDSARKELOKI	1005
Qy	1238	DDVIIIPFGESE-----EDYDDLQ--QVVTGE-----AVTPSVIDNLSKINIEYEV	1283
Db	1006	HEV--BEILNELOPNARALRYDEAGREVFENNTEOLKAEEKKILNOFTKIKKKRKE	1062

```

Qy 1284 LYKP-----LAGYRSLSKQLQENNVWTFVNVVKDILNRSFNKNRFKNVLESDLIP- 1333
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1063 LFEKTEDYVDHLDATYRELTKNPNSVNELAGGNASLRTIE---DEDEPFNAGIKYHATPP 1119
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 1336 ---YKDLTSSNYVVKDPYKFLNKRR-----DKFLSSSYNIKDSITDIDNFAND 1381
   ||:      :::::      :::::      :::::      :::::      :::::      :::::
Db 1120 LRFKDM-----EYLSGEKTVAAALLFAINSYQSPFFVLDEDAALDTN- 1167
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 1382 VLGYGKILSEKYKSDLSIKKKYNQKGNEKYLPFLNNIETLYKTVMNDKIDLIVHLEA 1441
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1168 -----VORIAAVIRRHNPDLQFI-----VISLNTMTEKSDALV----- 1202
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 1442 KVLNYTYEKSNVVEKIKELNY 1462
   | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1203 GYVRQQOENSCKIITLDSLNY 1223
   | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 40
RA50_AQUAE STANDARD; PRT; 978 AA..
ID RA50_AQUAE
AC O67124;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable DNA double-strand break repair rad50 Atphase.
GN RAD50 OR AQ.1006.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
NCBI_TAXID=63363;
RX NCBI
RN [1]
RC SEQUENCE FROM N.A.
SC STRAIN=VF5;
EX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., R.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Ausley M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mreII by unwinding
CC and/or repositioning DNA ends into the mreII active site (By
CC similarity).
CC CC -1- SUBUNIT: Forms a complex with mreII (By similarity).
CC CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commer-
CC entities requires a license agreement (See http://www.isb-sib.ch/annou-
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL; AE000718; AAC07092.1; -.
CC DR InterPro; IPR003439; ABC_transprot.
CC DR InterPro; IPR004592; SbcC.
CC DR InterPro; IPR002017; Spectrin.
CC DR TIGRfams; TIGR00618; sbcc_1.
CC KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 160 826 COILED COIL (POTENTIAL).
SQ SEQUENCE 978 AA; 115897 MW; 9B0F2BF51ADDD1151 CRC64;

Query Match 4.1%; Score 344.5; DB 1; Length 978;
Best Local Similarity 19.2%; Pred. No. 9e-05;
Matches 228; Conservative 232; Mismatches 353; Indels 377; Gaps

Qy 31 KLEALEDAVLTYSLPKRKMYLNGETSGTAVTTTPGSKGVASGGSGSVASGVS 90
   ||| | : | : : : | : : : | : | : | : | : | : | : | : | : | : | : | : |

```


Db 5 KLE-----VKFTYKKPQVI-----DFTF-LKFFVIQKGTG---AGKTSIID 43
Qy .91 GGSVASGSGSGSRRTPNSDSSDAKSYADLKHVRNVLITIKELKYPQ--- 146
Db 44 AITVALYGVPRYGA--SVATKYVLSRGEKELKVALDFSLGRNRYRVERIYREFPEDSOV 101
Qy 147 -LFDLTNHLMTLCDNIHGFYKLYDIDG--YEETINELLYKLNFFYDOLLRAKLNVDVANCYQI 203
Db 102 RVEEGRRLNIKANEVEKWLKPSGLDYKTFKTVILLPQGEFDFLKE-----SSERKKI 156
Qy 204 PFNLKIRANELDVLKLVGFRKPLDNKDNVGK-----MEDYIKNK-----KTIENINEL 255
Db 157 LINE-LGLELEKVRQLASFTFKNLEKREALKREYELLKDYTPTRKEVELEKTLKNLEE 215
Qy 256 IEESKKTIDKNKNTKEEKKKLQAOVDLSYKQLEEAHNLISVLEKRTDILKKN--- 312
Db 216 LKELTEEEKLROELKKAEEKD--SLRELQSVYTKLELEN-----LEKEVEKLEKLEF 269
Qy 313 -----ENTKELDLKINEIKNPPPPANGTNPNTLTKNKKIEEHEKEIKEIAKTI 361
Db 270 SRKVAPVPIAKRIEEDDKLTELK-----VRKNKL-----TKELAVLKDEL 311
Qy 362 KFNIDSLFTDPLEYLYREKNKNIDISAKVETKESTEPNPGVTPYPLSYNDINNALN 421
Db 312 SFAQELNRIEAEKEKEKEKEKELEHRLK-----KLOEIKEIKL 353
Qy 422 ELNSFGDLINPFDYTKPSKNIYTD-NER-----KKFNEIKEIT-KIEKKKIESDKKSYE 475
Db 354 ELSQSSSLKEKEREYQAKQEFEDLSERVEKGLKVAETEKELEKELKEFSEEEYSLK 413
Qy 476 DRKSLLNDITKEYEKLKLEIYDYSKFNNNIDLTNPEKMGKRYSKYKVEKTLHNTFASYEN 535
Db 414 MKERLLVELQRLKEL-----KEKEQOLENLQ-----KYKE 445
Qy 536 SKHNLKLTALKALMEDYSLRNIVVEKELKYKLNLSKIENEIT-----LVEN 584
Db 446 KKKVHEKVLNLEKE-----RELKERLHYHAHVAASYLSPGDTCPVCGGIYRGKALEN 500
Qy 585 IKKDEQLFEKKITKDENKPKDEKILEVSDIVKVOVKV-----LLMKNIDELKKTOL 636
Db 501 V--DAEGISELKHAKELKEER--EIDTTLKLYAQKINSKEMEKURNEVEELR--- 553
Qy 637 ILKNVELKHNTHVPSNYSKOENKQEPYVILVLKDEIKDLKVPMPKVESLINEKKNIKTEG 696
Db 554 -----EIPENLKER-----IKKLEELRIEKELEHKLKNYK----- 585
Qy 697 QSDNSEPTEGTEIGQATTGKQAGSALGDSVQAOQOQAPPPVPPVPEAKAQP 756
Db 586 -----ALE-----DRQKQK-----BAQAKL- 601
Qy 757 TPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHKYLVSHTMNEKILKQYKITKEESK 816
Db 602 -----HKAQTELEL-----LKEKREKSRLVKEPKE- 627
Qy 817 LSSCDPLDLQNIQNNIPVMYSMFDSNLNSLSQLFMETIYERKMCNLYKLKD-----NDK 871
Db 628 -----LYRVER-----LEDYEE-----SLKEEINYINSK 651
Qy 872 IKNLEPAKVVSTVKTLSSSSMQPLSLTPQDKPEVGSANDTSTNLSNLSKLFE--NI 929
Db 652 LQETEEREKKURKHFEELSSRSKSL-----EGELSALNESINSLESEERKEKLKELANI 704
Qy 930 LSLGKN-KNIIQELIGOKSSNFYEKILKQSDTFYNESFTNFVKSADDSINLSNDESKRK 988
Db 705 YEVAKSPEVVVELYLGQKAE--LERKIKE-----FEESFQSL-----KLKKS 745
Qy 989 KLEEDINKLTKLQSLFDLYNKKYKLEKLERLFDKPKTKYKQWQIKLTLKEQLESKLNS 1048
Db 746 EIEEKLKEYEIGRELS-DIKEYESVKTOLKEKHKLGEVK---RELEHLGERLKR--- 798
Qy 1049 LNNPKHVLQNFVFPFNKKKAEIAETENTLTKILKHKYKG-LVKYNGSGSSPLKTL-- 1105
Db 799 -----BELQKEISELEKKLEVYRVISNDFRGDRFQKYVSEIMLQKVYDR 842

Qy 1106 -SEESIOTEDNYASLENFKVLKLEGLKDKNLNLE-----KKKLSYLSGG 1149
Db 843 ASEFYFKTGNFY-----FELERATKGRDKDIVVESSTSQRRPVSSLSGG 888

RESULT 41

MYH7_MESAU STANDARD; PRT; 1934 AA.
ID MYH7_MESAU STANDARD; PRT; 1934 AA.
AC P13540; Q60540;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
GN MYH7.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FIB; TISSUE=Liver;
RX MEDLINE=95115033; PubMed=7815459;
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
heavy chain gene from Syrian hamster";
RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OF 962-1935 FROM N.A.
RX MEDLINE=88247788; PubMed=3380703;
RA Jandreski M.A., Sole M.J., Liew C.-C.;
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin
heavy chain";
RL Nucleic Acids Res. 16:4737-4737(1988).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEATITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; L12104; AAA62313.1; -;
CC EMBL; X07273; CAA30256.1; -;
CC PIR; A28298; A28298.
CC HSP; P08799; 1MND.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF01576; Myosin_tail; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS0096; IO; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 779 MYOSIN HEAD-LIKE.
FT 780 809 IQ.
FT DOMAIN 839 1934 COILED COIL (POTENTIAL).
FT NP_BIND 177 184 ATP.
FT DOMAIN 554 676 ACTIN-BINDING.
FT 756 770 ACTIN-BINDING.
FT MOD_RES 128 128 METHYLATION (SH-1) (POTENTIAL).
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 966 966 D -> E (IN REF. 2).
FT CONFLICT 978 978 T -> TE (IN REF. 2).
FT CONFLICT 986 986 E -> Q (IN REF. 2).
FT CONFLICT 1008 1014 DLQAED -> ALEAKT (IN REF. 2).
FT CONFLICT 1057 1057 D -> Y (IN REF. 2).
FT CONFLICT 1060 1060 L -> V (IN REF. 2).
FT CONFLICT 1095 1095 D -> N (IN REF. 2).
FT CONFLICT 1217 1217 E -> D (IN REF. 2).
FT CONFLICT 1271 1271 D -> N (IN REF. 2).
FT CONFLICT 1327 1327 T -> A (IN REF. 2).
FT CONFLICT 1358 1358 C -> R (IN REF. 2).
FT CONFLICT 1504 1504 L -> V (IN REF. 2).
FT CONFLICT 1537 1537 M -> L (IN REF. 2).
FT CONFLICT 1556 1556 N -> K (IN REF. 2).
SQ SEQUENCE 1934 AA; 222928 MW; FDBAC59310B0B57D CRC64;

Query Match 4.1%; Score 344.5; DB 1; Length 1934;
Best Local Similarity 19.4%; Pred. No. 0.0019;
Matches 290; Conservative 259; Mismatches 593; Indels 353; Gaps 62;

QY 47 QKEKVLNKGSTGATVTTTPGSKGVSAS-GSGGVSASGSGVSASGSGVSASGGS 105
DB 594 QNKDPLNVTVGLYQKSLKLLNLFANYAGADAPVDKGGKAKKSGSFQTVLHREN 653
QY 106 GNSRRTNPDSNDSDAKYADLKRVNRYLLTIKELYPQLFD--LTHNMLTLCDNING 163
DB 654 LNKMLNLRSTH-----PHEVR--CIIPNETKSPGVMDNPLVMHQLRCNGVLGS 700
QY 164 FYLIDGYEINEELLYK-----LN-----FYFDLLRA--KLNDVCANDYQOIPN 206
DB 701 IRICRKGFP--NRILYDGFQRYLINLPAAPGOFIDSRKGAELSLSLDIDHNQYKFG 758
QY 207 -----LKRANELVDLKKLVFGYRKPLD-----N 230
DB 759 HTKVFFKAGLLGLEEMRDLRSRIITRIQASRGLLSRMEF--KKLLERRDLSLLVIQWN 816
QY 231 IKDNVG-----KMEYIK--KNKKTIENTINELIEESKKTIDKNKNATKEEEKK- 276
DB 817 IRAFMGVKNWPMKLYFKIKPLKSAETEKEMATKKEEFGRYKDALEKSEARRKELEEK 876
QY 277 -KLYQAOYDLSTYNKOLEFAHNLSIVLEKRDITLKNK-----NIKELLDKINEKNPPP 330
DB 877 VSLLOEKNDQL---QVQAEQDNLADEERCQDLKKNKIQLEAKVKEMTERLEDEE---- 929
QY 331 ANSGNTPNTLLDNKKKIEEKEIKIAKTIFKNIDSLFTDPLELYLYLREKNKNIDISA 390
DB 930 -----EMNAELTAKKRLEDECESELKRDID-----DLELTL-----A 961
QY 391 KVE-TKESTPEPNVGVTYPLSYNDINNALNELNLSFGDLINPFDYTKPESKNIVTDNER 449
DB 962 KVEKDKHATE-----NKKVKNLTTEAGLDEIIAKL--TKE----- 994
QY 450 KKFINEIKKIKIEKKIESDKSYEDRSKLSINDITKEYEKLLNEIYDSKFNNNIDLTNF 509
DB 995 KRALQEAHQO-ALDDILQAEEDK-----VNTLTFSKVKLEQQVDD--LEGSLQEKK 1042
QY 510 EKMGMKRYSKYE---KLTHHTTFASYENSKHNL-EKLTK-----ALK-YMEDYSLRWI 558
DB 1043 VRMDLERAKRKLEGLDKLT-QESIMDLENKQOLDEKLLKKDFELNALNARIEDEQALGS 1101

QY 559 VVEKELKYKKNLISKIENIEITLVENIKKDEQLFEKKITKDKENKPKDEKILEVSDIVKVO 618
DB 1102 QLOKKLKELOARIBELEBELE--AERTAKVKEKLRSDLSRELEISERLEAGGATSVQ 1159
QY 619 VOKVLLMNK--IDELKTKTQLILKYNVELKHNIHVPSYKQENKQEPYLYLVLKKEIDKLKV 676
DB 1160 IE---MNNKREAFQKRRDLEEAATLQHEA---TAAALRKKHADSAELGEQIDNLQ- 1210
QY 677 FMPKVESLINEEKKIKTEGOSDNSEPTGEITQAT-----TKPQOQASALEGDSV 730
DB 1211 ---RVQKLEKESEFKLE--LDDVTSNMEQIIKAKANLEKMCRTLEDOMNHRSAEET 1265
QY 731 QAQAQEQKQAQPPVPVPPVPEAKAQAQPTPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHK 790
DB 1266 QRSVNDLTSORAKLOTENGELSRLDKEALISQLTR--GKLTYYTOOLEDL----- 1314
QY 791 YILVSHSTWNEKILKQYKITKEESKLSKSCDPLDLLFNQNNIPVMSFDSLNNSLSOL 850
DB 1315 -----KQLEEEVKAKNTLAHALQSAHRHCDLLREQYEETEAKABLCQVLSKANSEVAQ 1369
QY 851 FMEIYEKEMVCNLYKLD--NDKIKNLLKEAKKYSTVSKTSSSMOPLSLTPQDKRPE--- 906
DB 1370 WRTYETDAIQTETEELAEAKKLAORLQDAEEAVNA--KCSSLEKTKHRLONEIEDLM 1428
QY 907 VSANDDTSHSTNLNLSKLFINLSLGNKNKIYQELIGOKSS-----ENF 951
DB 1429 VDVERSNAAAALDKQRNFDKILAEWKQ--YEESQSELESSQKEARSLSLTFELKKN 1486
QY 952 YEKILKDSDTFYNESEFTNVKSKADDSLND-----ESKRKKLEEDINKLAK 999
DB 1487 YEESEHLEHTFKREN-----KNLQEEISDLTEQLGSTGKSIHEIKIRKQAEKMELOS 1541
QY 1000 TL-----OLSFDLNLYKYLKLER-LFDKKTVTGKYKM-QIKKLTLLKE 1040
DB 1542 ALAEAASLEHEEGNILLRAQLEF---NQIKAEIERKLAEKDEMEQAKRNHLRVYDSLQT 1598
QY 1041 QLESKLSLNNPKHVLQNFVFFNKKKEAEIAETENTLTKILLKHYKGLVYKYNNGESS 1100
DB 1599 SUDAETRSRNEALRV-----KKMEGDLNEMEIQLSHANRMAAEQAQVK-----S 1644
QY 1101 PLKLTISEESTOTEDNVASLENFKVLSKLGKLDNLNLEKKKLSYSSSGLHLHLIAELKEV 1160
DB 1645 LOSLLKDTQIOLDDAVRANDD-----LKENAIVERNNLLQAELEELRAVVEQT 1694
QY 1161 IKNNYVTGNSPENNTDV-----NNALESYKFLF-PEGTDVATVYVSESGSDTLEQSPK 1213
DB 1695 ERSKRLAEQELIETSERVQLLHSHQNTSLNQKKMDADLSQLQTEVEEAQVQECRNEAKA 1754
QY 1214 KPASTHVGAESNTITTSQ-----NVDDEVDDVIIVPIFGSEEDYDDLQGV-VT 1261
DB 1755 KKAITDAAMMAEELKEQDTSAHLERMKKNMEQTKDL-----QHRLDEAEQIALK 1805
QY 1262 GEAVTPSVINDILSKTENEVEVLYLKLPLAGVYRSKLKOLENNVMTFNVAVKDILNSRENK 1321
DB 1806 GKQKQLQKLEARVRELENELE-----AEQKRNAESVKGMKRSERIKELTYOT 1853
QY 1322 RENFNKVLSDLPYKDLTSSNVVVKDPYKFLNKEKRRKF---LSSYNYIKDSID 1373
DB 1854 EEDRNLLR-----LQDLVDKLQLKVKYKQAEAEAEQANTNLSKFRVQVHOLD 1903

RESULT 42
YCF2_MARPO
ID YCF2_MARPO STANDARD; PRT; 2136 AA.
AC P09375;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 259 kDa protein ycf2 (ORF 2136).
GN YCF2
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
CC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;
CC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN
RP SEQUENCE FROM N.A.
RA Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.;
RT Liverwort Marchantia polymorpha chloroplast DNA.;
RL Nature 322:572-574(1986).
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X04465; CAA28078.1; .
DR PIR; A05037; A05037.
DR PIR; S01591; S01591.
DR InterPro; IPR003959; AAA_ATPase_centra.
DR Pfam; PF00004; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match 4.1%; Score 344; DB 1; Length 2136;
Best Local Similarity 19.8%; Pred. No. 0.00024;
Matches 338; Conservative 243; Mismatches 541; Indels 584; Gaps 80;

Qy 160 NIHGFKYLDIGYEINELLYKLN---FYF-----DILRAKLDN-----VC 196
Db 204 NIDELKHFEEY--INQYIPNNWMEYFYIFINQLKIDIKNSKNGSGFVFLAFC 261
Qy 197 ANDYQIPNFKIRANELDVLKLVFGYKRPDLNIDKNVGMEDYI---KK---NKKTI 249
Db 262 EKLLEFEVFLSPNNNNLQ--KL-----NCLNFSFELDFICILNKKLPWVKKIF 310
Qy 250 ENINELIEESKTTIDKNKNATKEEKKLYQAOYDLSIYNNKOLBEAHLNISVLEKRIDL 309
Db 311 KNLQNFESDRKLLIE-----SFFLLKIKGNLYFKNYTEFVTWQSYKKDCLDFN 358
Qy 310 KKNE-NIKELLDKINEIKNPPAN-----SGNTPNTPLLDKNKKIEHEKEIETKI 361
Db 359 KFNELNNEIYKIEELFSDIYKFSKYLYEGKKSKRTIHKOSFNNNIYYKKLSI---- 414
Qy 362 KFNIDSLF-----TDPLELYYLREK-----NKN 385
Db 415 -FNFTIPYDNNLLFDWLAKKNYINNPKFLSLIYSSISNQIFLFFKQKNSKFNKN 473
Qy 396 I-----DISAKVETKS-TEPNYPNGVTYPLSYNDINNALNELNSGDLINPDYTK 438
Db 474 LVKNSKDVITNVFSKENKIEINNFSKIY-----AFFELISNEIDNKFVINKI 524
Qy 439 PSKNLYTDNERKKFNEIKER-----IKIEKKIESDKKSYEDRSKSLN---DITKE--Y 488
Db 525 SLKNINKKQKRFYLNKIKSSDNFRFINLWKIKYSSQOFVSNNSFLNPAFELQQNY 584
Qy 489 EKLNEIYDSSFNNNIDLTNEKMGKRYSYKVEKLTHTNPFASYEN--SKHNLEKLTKA 546
Db 595 LKKKNILFFKKLINE-----VPSNFEYFOY-YCKKLNIFLAFASLEKILKARN-KKFTIS 637
Qy 547 LKYMEDYSLRNVVKEKLYKYNLISKITENEIETLVENIKKDEQOLFPEKTKIDENKPE 606
Db 638 IKLEK-----KFKYKNLN--ENGEYKIESQILQNEKELNKKRKNQFQNPNI 682
Qy 607 KILEVSDIVKQV-----QKVLMMNKIDELKK--TOLLKKNVEL-----KHNHVP 650
Db 683 KILSFYSSKKNIYLNQYFFKNKLNINNKLTITWKIKSNKLVISSEYNNKIIWNKKMKFF 742

Qy 651 NSYKOE-----NKQEPYIYLIVLKKKEIDKLVKFMKPVESLIN-----EEKNKITEGOS 698
Db 743 SFSKNSVLDTFFFNKKSFNILITVIFDKLKKIQLNFOEIQILNCFSLFNSKNIKTKIF 802
Qy 699 DNSEPTSEIGTQATTKPGQOAGSALBGDSVQAOAQOQOAPPPVPVPEAKAQVPTP 758
Db 803 KNSYFINEN-----LTT 814
Qy 759 PAPVNNKTENVSKLDYLEKLYEFLNTSYICH---KYILVSHSTMNEKILKQYKITEEES 815
Db 815 TFSFNDKEFNIF---FLELFISEINNDFLMRFKKYL-----YRIYKDEI 858
Qy 816 KLSSCDPIDLDFNIGNNIPVMSFDSLNN---SLSQLFMEIYEREMVCN---LYK--- 865
Db 859 LFNPIENQOLLQNFPEKTKIL-TFIDFLQDPELVNNRFIFLHEKTKIKNNLLYLRLK 917
Qy 866 --LKDN-----DKTNLLEBAKVVSTVKTLSSSSMOPLSLTPQDKPEVSANDTSHST 917
Db 918 IFLDKRNFLLINEIKSFIE--KKNLFIKLSQLSNVL----- 953
Qy 918 NLNLSKLIFENLTL---GKNNIYQELIGOKSSENYEK----- 954
Db 954 -VKNSYKFFDNIFNPHFLKQKEKNI--EII--LNNQNYFEKSLAKTYLKNLNLNNSYSK 1008
Qy 955 -----ILKDS-----TFYNESFTNVFVSKADDIN---SLNDESKRKK 989
Db 1009 FSYKIFIFOLLNLNKNKYKTFOWISELIFYSKNLNYKIQNKIENNNYCNKNIYKXKK 1068
Qy 990 LEE-----DINKLKT-----LQSLFDLYNKYL-----KLERLEFDKKTGVK- 1027
Db 1069 IKTWNFEKNNLFQTNNSWFTLEWMEYNTVILLQIQETFFQITDVLVEYFKKKIIEKN 1128
Qy 1028 -----YKMOIKLTLKEQLESL---NSLNNPKHVLPNSFFFNKKKEAIEAETENT 1077
Db 1129 LKFFLKSKKISLKTLSFHNFKLNMNLRFFNEINYKKNLLNF-----LWSDFN 1177
Qy 1078 LENTKILLKHYKGLV---KYVNGESSPLKLTSESIQTEONYASLENPKVL----- 1125
Db 1178 INNCCNLVWVLSLVIFIFLYQ-----KIFS---IIIGSCDFHLWKKFIEIIQYLTDRSR 1229
Qy 1126 ---SKLEGKLDNINLEKKKLSYSSGLHLHIAELKEVINKNKYTGNSPSNNTDVNNA 1181
Db 1230 SLYFTKLTRRNKALNKTENLLSYFQNLTHYITNIKFYLLTK----- 1272
Qy 1182 LESYKKLPEGTDVATVVSSESGDTLEOSQPKPASTHVGAESNTITTSQNVDDDEVDD-- 1239
Db 1273 -KNLKKWL-----INNKTLDJSRRKRLVLOSILTHNKI---ONYGFELSNK 1316
Qy 1240 VIIVPIFGESEBDDLGQVVTGEAVTPSVIDNIIKLSIENEYEVLYLKLPGVYSLKKQ 1299
Db 1317 QFTSYFG-----YOITNOOGLLYFOYLAOFF---QKN 1346
Qy 1300 LENNVN-----TFNVNVKDLNRSRKNREN-----FKNV---LESDLIPYKDL----- 1339
Db 1347 LINSLDLANKWIVFSEFWHKIFSOKLRQTNNIELGFQNPVLPQFGLSYSGKILLIGPI 1406
Qy 1340 -TSSNVYVKD-----PYKF-LNK-----EKROKFLSSY----- 1365
Db 1407 ETGRSYLIKLNLAESYVPLFKISINKLKNYPDVITESWMNIIIESLRNLNLTDFAKKM 1466
Qy 1366 -----NYIKDSIDTDINFANDVLYGYKILSEKYSKSDLSIRK---YINDK 1407
Db 1467 SPCLIIWQNIHQLNVNRLTQNVESDPTF---LLG---ILLKYFQDFSKTKNNIIVIGS 1520
Qy 1408 QGENEYLPFLNNIETLYKTVNDKIDLVHLEAKLVNITYEKSNNVEVK----- 1456
Db 1521 THLPKKVDPALISPNRLDKIIN--VRLENI SORKQKQFPLLLKKNFOLKNFLFNFEFGS 1578
Qy 1457 -----IKELNYLTKIQDKLADFKNNNFVGLADLSTQYNH-----NNLLTKFLSTG 1502
Db 1579 RTMGYNLRDL SAL-TNEVLLISITKNRSFIDTDTLKLAFHFRIGLTYTNNKLNFDRIFK 1637

QY 1503 MYFENLAKTVLSN-LLDGNLQGLMNI 1527
ID MYH6_HUMAN STANDARD; PRT; 1939 AA.
Db 1638 IVIYKVGKTIQNILIKSSMNLNI 1663

RESULT 43

MYH6_HUMAN
AC P13533; Q13943; Q14906; Q14907;
DT 01-JAN-1990 (Rel. 13, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
GN MYH6 OR MYHCA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92133665; PubMed=1776652;
RA Matsuoka R., Beisel K.W., Furutani M., Arai S., Takao A.;
RT "Complete sequence of human cardiac alpha-myosin heavy chain gene and
RT amino acid comparison to other myosins based on structural and
RT functional differences.";
RL Am. J. Med. Genet. 41:537-547(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94140346; PubMed=8307559;
RA Epp T.A., Dixon I.M., Wang H.Y., Sole M.J., Liew C.-C.;
RT "Structural organization of the human cardiac alpha-myosin heavy
RT chain gene (MYH6).";
RL Genomics 18:505-509(1993).
RN [3]
RP SEQUENCE OF 1-177 AND 1551-1939 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN [5]
RP SEQUENCE OF 1407-1939 FROM N.A.
RX MEDLINE=88299163; PubMed=2969919;
RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
RT "Molecular cloning and characterization of human cardiac alpha- and
RT beta-form myosin heavy chain complementary DNA clones. Regulation of
RT expression during development and pressure overload in human
RT atrium.";
RL J. Clin. Invest. 82:524-531(1988).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D00943; BAA00791.1; -;
DR EMBL; Z20656; CAA79675.1; -;
DR EMBL; M25140; AAA60386.1; -;
DR EMBL; M25162; AAA60386.1; JOINED.
DR EMBL; M25142; AAA60387.1; -;
DR EMBL; M25141; AAA60387.1; JOINED.
DR EMBL; M25141; AAA60387.1; JOINED.
DR PIR; A46762; A46762.
DR PIR; A28908; A28908.
DR HSSP; P08799; LMND.
DR Genew; HGNC:7576; MYH6.
DR MIM; 160710; -;
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 2; myosin_tail; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IQ.
FT DOMAIN 842 1939 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP.
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 88 88 Q -> E (IN REF. 1).
FT CONFLICT 574 574 Q -> P (IN REF. 1).
FT CONFLICT 608 608 A -> G (IN REF. 1).
FT CONFLICT 744 744 T -> A (IN REF. 1).
FT CONFLICT 790 790 M -> I (IN REF. 1).
FT CONFLICT 1014 1014 V -> A (IN REF. 1).
FT CONFLICT 1021 1021 S -> T (IN REF. 1).
FT CONFLICT 1101 1101 A -> V (IN REF. 1).
FT CONFLICT 1290 1290 A -> S (IN REF. 1).
FT CONFLICT 1373 1373 W -> C (IN REF. 1).
FT CONFLICT 1533 1533 K -> N (IN REF. 5).
FT CONFLICT 1540 1540 L -> M (IN REF. 5).
FT CONFLICT 1577 1578 KL -> NV (IN REF. 5).
FT CONFLICT 1705 1706 EQ -> DR (IN REF. 1).
FT CONFLICT 1733 1733 E -> D (IN REF. 1).
FT CONFLICT 1734 1734 A -> S (IN REF. 2).
FT CONFLICT 1737 1737 T -> S (IN REF. 1).
FT CONFLICT 1763 1763 D -> H (IN REF. 1).
FT CONFLICT 1788 1788 M -> I (IN REF. 3).
FT CONFLICT 1871 1871 D -> N (IN REF. 5).
FT CONFLICT 1882 1882 R -> G (IN REF. 5).
FT CONFLICT 1890 1890 Q -> R (IN REF. 5).
FT CONFLICT 1933 1933 MISSING (IN REF. 5).
SQ SEQUENCE 1939 AA; 223689 MW; ECB87E7CEB768B6F CRC64;

Query Match 4.1%; Score 343; DB 1; Length 1939;
Best Local Similarity 19.7%; Pred. No. 0.00021;
Matches 304; Conservative 267; Mismatches 542; Indels 432; Gaps 70;

QY 28 LVKLEALEDAVLTVSLFQKMKVNLNEGTSCTAVTTSTPGKSGSVASGSGSVASGGS 87

```
Db 595 LEKNDPLNETVW---ALVQSSSLK-----MATLFS-----SYATADTGDSKSGK 639
Qy 88 VASGSGVASGSGVASGSG-----NSRRTPNSDSSDAKSYADLKHVRNVLITIKELK 143
Db 640 KKGSGFQTVSALHRENKLNKLTNRTHHP-----HEVR---CIIPNERK 681
Qy 144 YPOLFD--LTHNMLPLCDNIHGFVKYLDIGYEEINELLYK-----LN-----EYFD 186
Db 682 APGVNDPLVHQLRCNGVLEIGIRCRGFP--NRILGDFRQRYRILNPVAPGQFID 739
Qy 187 LLRA--KINDVANDYCOIPF---NLKIRANELDLVKK-----L 220
Db 740 SRKGTEKLLSLIDHNGYKFGHTKVFVKAGLLGLEEMRDERLSRIITRMQAQARGQLM 799
Qy 221 VFGYRKPLD-----NIKDNV-----KWDYIK--KNKKTIENINELIE 257
Db 800 RIEFKKIIVERDALLIVIQNIRAFNGVKNWPMKLYFKIKPLLKSAETEKEMATMKKEFG 859
Qy 258 ESKKTIDKNKATBEEKK--KLYQAQYDLSTYNKOLEEAHNLSVLEKRIIDTLKKNB-- 313
Db 860 RIKETLEKSEARRKELEKMYSLLOEKNDLQ---QVQAQDNLDNDAERCQDLKNIQ 916
Qy 314 ---NIKELDKINEIKNPPPPANGTPTNTLLDKNKKIEBEHEKEIKEIAKTIKFNIDSFT 370
Db 917 LEAKVKEMNERLEDEE---MNA-----ELTAKKRKLED---ECSELKKDI----- 956
Qy 371 DPLELEYLREKNKIDISAKVE-TKESTEPNEYPNGVTYPLSYNDINNALNELNSFDGL 429
Db 957 DDLLEJTL-----AKVEKEKHATE-----NKVKNLTTEEMAGLDEI 990
Qy 430 INPFDTYKTSKNIYTDNERKKFINKIEKIKIEKKKIESDKSYEDRSKSLNDITKEVE 489
Db 991 TAKL--TKE-----KALQEAHQ--ALDDLQVEEDK-----VNSLSKSKV 1027
Qy 490 KLLNIYDSKFNNDLNFKEKMGKRYSYKVE---KLTHNTFASYENSKHNLEKLTKA 546
Db 1028 KLEQQVDD--LEGSLEQKEKVRMDLERAKRKEGLDKLT-QESIMDLENDLQLEBEKLLK 1084
Qy 547 LKY-----MEDYSLRNIVVEKELKYKNLSKIENIEIETLVENIKKDEEOLFEKIT 598
Db 1085 KEFDINQNSKIEDQALQLOKLLKENQARIEELEELE--AERTARAKVEKURSDLS 1142
Qy 599 KDNKPDEKILEVSDIVKVQVKVLLMNK--IDELKTKQLIKNVKELKHNIHVPSYKQE 656
Db 1143 RELEBISERLEEAGGATSVQIE---MKNKREAEQKMRDLLEATLOHEA---TAAL 1194
Qy 657 NKQEPYILIVLKEIDKLKVPKVESLINEKKNIKTEGOSDSEPTGEITGOATTK 716
Db 1195 RKKHADSVAELEGEIDNLQ---RVQKLEKESEFKLE--LDDVTSNMEQIIKAKANLE 1248
Qy 717 PQQAAGSALGDSVQAQAEQKQAPPPVPVPEAKAQQV---PTPPAPVNNKTENVSKLD 773
Db 1249 KVSF-----TLDDQANETV-----VKLEEAQRLSDFTTORAKL--OTENGELAR 1291
Qy 774 YLEKLYEFLNTSYICHKYLIVSHSTMNE---KILQYKTKTEESKLSL----- 819
Db 1292 QLEE-----KEALISQTRGLSYTQQMEDIKQLEEBGKAKNALAHALQSA 1339
Qy 820 --CDPLDLFLNIQNIIPWYMSFDSLNNSLSOLFMEIYEKVMCLYKLD-NDKIKNLL 876
Db 1340 HDCDLLRQYEEETEAKAELQRLVLSKANSEVAQWRTKYETAQIORTTEELEAKKLAQRL 1399
Qy 877 EEAKKVSIVKTVLSSSSQPLSLTPQDKPE---VSANDDTSHSTNLSKLFENILSLG 933
Db 1400 QDAEEAVEAVNA-KCSSLEKTKHRLQNEIEDLMDVDSERNAALAKQRNFKILAEW 1458
Qy 934 KNKNITQELIGOKSS-----ENFYEKILKDSFTFYNESFTNFVKSKADDI 978
Db 1459 KQK--YEESQSELESSQKARSLSLTFELFKNAYEESLEHLETFKREN-----KNLQEI 1511
Qy 979 NSLND-----ESKRKKLEEDINKLTKTL-----QLSFDLYN 1009
```

```
Db 1512 SDLTEQLGEGGKNVHELEKVRKQLEVEKLEQLSALAEAEASLEHEEGKITLRAQLEF---N 1568
Qy 1010 KYKLLKLER-LFDKKTVTGKYKMQIKKLT-LLEKEQLSKLNSLNNPKHYLVQNSVFENKKK 1067
Db 1569 QIKABIERKLAEKDEMEQAKRNHORVVDLSQTSLSAETRSRNEVLRV-----KKKM 1620
Qy 1068 EAEIAETENTLENTKILLKHYKGLVYNGESSPLTKTSEESIQTEDNYASLENFKVLSK 1127
Db 1621 EGDNLNEMEIQLSHANPMAAEAKQVK-----SLOSLLKDTQIOLDADVANDD----- 1668
Qy 1128 LEGKLDKNLEKLLSYLSGLHLHIAEL-----KEVINKNYTGNSPENNT 1176
Db 1669 ---LKENIAIVERNNLLQAELEELRAVVEQTERSRLAEQELIETSERVOLLHSQNTS 1724
Qy 1177 DVN-----NALESYKFKLPDGTDTATVVSSESGSDTLEOSOP 1212
Db 1725 LINQKKKWEADLTQLOSEVEAEVQECRNAEAKKAI---TDAAMMA-----EELKK 1773
Qy 1213 KKPASTHV-GAESNTITTSQNVDDVDVVIIPFGESEEDYDLGQVVVTGEAVTPSVID 1271
Db 1774 EODTSAHLERMKNKMEQTIKDQHLRDLDAEQIALAGGKKQ----- 1813
Qy 1272 NILSKIENEYEVLYLKLPLAGVYRSILKQLENNVMTFNVNVKDIILNSRFRNKNFNKVL 1331
Db 1814 --LQKLEAR-----VRELEGELEAEQKRNAESVGMKRSERRIKELTYQTEEDKKNLLR- 1865
Qy 1332 DLIPYKDLTSSNVYVVDYKPKFLNKEKROKF---LSSYNYKIDSID 1373
Db 1866 ---LQDLVDLQLKVKAYKQAEAEBOANTLSKFRKRVQHELD 1906

RESULT 44
MYH6_MOUSE
ID MYH6_MOUSE STANDARD; PRT; 1938 AA.
AC Q02566; Q64258; Q64738;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
GN MYH6 OR MYHCA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, A/J, C57BL/6J, and DBA/2J;
RX MEDLINE=92250040; PubMed=1577481;
RA Quinn-Laquer B.K., Kennedy J.E., Wei S.J., Beisel K.W.;
RT "Characterization of the allelic differences in the mouse cardiac
alpha-myosin heavy chain coding sequence.";
RL Genomics 13:176-188(1992).
RN [2]
RP SEQUENCE OF 1-67 FROM N.A.
RC STRAIN=AKR;
RX MEDLINE=91225025; PubMed=2026617;
RA Gulick J., Subramaniam A., Neumann J., Robbins J.;
RT "Isolation and characterization of the mouse cardiac myosin heavy
chain genes.";
RL J. Biol. Chem. 266:9180-9185(1991).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
```

CC SUBFRAGMENT (S2).

CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE

CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.

CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M76598; AAA37159.1; -

DR EMBL; M76599; AAA37160.1; -

DR EMBL; M76600; AAA37161.1; -

DR EMBL; M76601; AAA37162.1; -

DR EMBL; M62404; AAA37424.1; -

DR HSSP; P08799; IMND.

DR SWISS-2DPAGE; Q02566; MOUSE.

DR MGD; MGI:97255; Myhca.

DR InterPro; IPR000048; IQ_region.

DR InterPro; IPR004009; Myosin_N.

DR InterPro; IPR002928; Myosin_tail.

DR InterPro; IPR001609; myosin_head.

DR Pfam; PF00612; IQ; 2.

DR Pfam; PF01576; Myosin_tail; 1.

DR Pfam; PF02736; Myosin_N; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR ProDom; PD000355; myosin_head; 1.

DR SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.

DR MYOSIN; Muscle protein; Coiled coil; Thick filament; Actin-binding;

KW ATP-binding; Methylation; Alkylation; Multigene family;

KW Calmodulin-binding; Polymorphism.

FT DOMAIN 1 782 MYOSIN HEAD-LIKE.

FT DOMAIN 783 812 IQ.

FT DOMAIN 842 1938 COILED COIL (POTENTIAL).

FT NP_BIND 178 185 ATP.

FT DOMAIN 657 679 ACTIN-BINDING.

FT DOMAIN 759 773 ACTIN-BINDING.

FT DOMAIN 790 807 CALMODULIN-BINDING (BY SIMILARITY).

FT DOMAIN 816 833 CALMODULIN-BINDING (BY SIMILARITY).

FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).

FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).

FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).

FT VARIANT 194 194 Y -> D.

FT VARIANT 545 545 S -> A.

FT VARIANT 838 838 I -> S.

SQ SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;

Query Match 4.1%; Score 341.5; DB 1; Length 1938;

Best Local Similarity 20.0%; Pred. No. 0.00024;

Matches 251; Conservative 234; Mismatches 467; Indels 303; Gaps 56;

QY 237 KMEDYK--KNKKTENINELTESKTTIDKNKATKEEKK--KLYQAQYDLISYINKOL 292

DB 837 KRLPLKSAETKEANMKKEGRVQDALEKSEARKEEKEKWSLLQEKNDLQL---QV 893

QY 293 EBAHNLISVLEKRIDTLTKNE-----NIKELDKINEIKNPPPPANSNGNPTNTLLDKNKKI 347

DB 894 QAEQNDLDAEERCQOLIKNKIQLEAKYKEMTERLEDEEE--MNA-----ELTAKKRKL 945

QY 348 EHEKEIKIATIKFNDSLTDFLELYLREKNKNKNDISAKVE-TKESTEPNYPNG 406

DB 946 ED---ECSELKKDI-----DQLETL-----AKVEKEKHATE----- 974

QY 407 VTYPLSYNDINNALNELNSFGDLPNFDYTKPSKNIYTDNERKKFINKIEKIKIEKK 466

DB 975 -----NKVKNLTEEMAGLDEIAKL--TKE-----KRALQEAHQO--ALDDLQ 1013

QY 467 IESDKSYEDRSKSLNDITKEYEKLLEIYDSKFNNNIDLTNPFKMMGRKRYKYVE---K 523

DB 1014 AEEK---VNTLTSKVKLEQQVDD--LEGSLQEKKKVRMDLERAKRKLEGLK 1062

QY 524 LTHHTFASYENSKHLEKLTALKY-----MEDYSLRNIVVEKLYKYNLISKIE 575

DB 1063 LT-QESIMDLNDKLEKKKKKEFDISQQNSKIEDEQALALQLOKLEKINQARLEE 1121

QY 576 NEIETLVENIKKDEQLPEKTKTDENKPKDEKILEVSDIVKQVQKVLMMNK--IDELKK 633

DB 1122 EELE--AERTARAKVEKRLSDLSRELEISERLEEAGGATSVQIE---MNRKREAFQ 1175

QY 634 TQLILKNVELKINIHPVNSYKQENKQEPYILVLKLEIDKLVFMPKVESLINEEKNK 693

DB 1176 MRDLEEAATLQHEA---TAAALRKKHADSVAGELGQIDNLQ---RVQKLEKEKSEF 1227

QY 694 TEGQSDNSEPSTEGETGQATTKPGQAGSALGEGSVQAQAEQKQAPVPVPVPEAK 753

DB 1228 LE--LDDVTSNMEQITIKAKANLEKYSR-----TLEQANEYR-----VKLEAQ 1270

QY 754 QV---PTPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHYILVSHSTMNE---KILQ 806

DB 1271 SLNDETTQRAKL--QTENGELARQLEE-----KEALISQLTGKLSYTOQMED 1316

QY 807 YKITREESKLSS-----CDPLDLLFNQNNIPVMYMPDLSLNNLSQLFMEI 854

DB 1317 LKQLEEEGKAKNALAHALQSSRHDCDLLREQVEEEMEAQELQRLVLSKANSEVAQWRTK 1376

QY 855 YEKEMVCNLYKLD-NDKIKNLLEBAKKYSTSVKTLSSSSMQPLSLTPQDKPE---VSN 910

DB 1377 YETDAIQRTTELEEAQKKLAQLODAEAEAVNA-KCSSLEKTKHRLQNEIEDLMVDVE 1435

QY 911 DDTSHSTNLNLSIKLFENILSLGKNKIYOELIGKSS-----ENFYEKI 955

DB 1436 RNAAAAALDKQRFDKILAEWKQ--YEESQSELESSQKEARSSTLFLKKNAYES 1493

QY 956 LKSDTFYNESTFNFVSKADDINSND-----ESKRKKLEEDINKKTL-- 1001

DB 1494 LEHLETFKREN---KNLQEBISDLTEQLEGEGKNVHELEKIRKQLEVEKLEQALAE 1548

QY 1002 -----QLSFDLYNKYKLEKLER-LFDKKTGVGKYM-QIKKLTLLKQLES 1044

DB 1549 AEASLEHEGKILRAQLEF---NQIKAEIERKLAEDDEMEQAKRNHLRWVDSLQSLDA 1605

QY 1045 KLSLNNPKHVLQNFVFPNKKKEAEIAETENTLENTKILL---KHYKGLVKYYNGESS 1100

DB 1606 ETRSRNEALRV-----KKMEGDLNEMEIQLSOANRIASEAQKHLK-----NSQAH 1651

QY 1101 PLKTLSEESIQTEDNYASLENFVKLSKLEKLDNLEKKNLSYSSGLHHLIAELKEV 1160

DB 1652 ---LKDQLOLDADVHANDD-----LKENIAIVERRRNLLQAELEELRAYVEOT 1697

QY 1161 IKKNYTGNSPSENNTDV-----NNALESYKFKFPEG-TDVATVYSESGSDTLEQSOPK 1213

DB 1698 ESKRLAEQELLETSEKRVOLLHSQNTSLNQKKKMSDLSLTQLOTEVEEAQVQECRNEAEKA 1757

QY 1214 KPASTHVGAESNTITTSQ-----NVDDEVDVDTIIVPIFGSESEDDYDLQGV-VT 1261

DB 1758 KKAITDAAMWAEELKEQDTSALHERMKNMEQT IKDL-----OHRLEAEQIALK 1808

QY 1262 GBAVTPSVTDNTLSKIENYEYVLYLKLPLAGVYRSKLQLENNVMTNVNKOILNSRFNK 1321

DB 1809 GKKQLQKLEARVRELENELE-----AEQKRAESVKGMRKSRRIKELTYQT 1856

QY 1322 RENFNKVLSDILPYKDLTSSNVVVKDPYKFLNKEKDKF---LSSYNYIKDSID 1373

DB 1857 EEDKN-----LMRLQDLVDKQLQKLVKAYKQAEAEAEQANLNSKFRKVQHELD 1906

RESULT 45

MYHB_CHKCK

ID MYHB_CHKCK

STANDARD;

PRT; 1978 AA.


```
Db 1393 -----ETMEGKKLQRIEISTQOFEKAASYDKLETKN-----RLQOE 1433
QY 847 LSOLFMEI-YEKEMVNLVYKLDNDKIKNLLLEAKVST-----SVKTL 889
Db 1434 LDDLVDLQDQOLVSNLE--KKQKFDQMLAEKNISSKYADERDRAEAREKETRAL 1491
QY 890 SSSSMQPLSITPDQKPE-----VSGANDTSHSTN-----LNN 921
Db 1492 SLARALEEALEKEELERTNKMKAEMEDLVSKDDVGNVHELEKSKRTLSQOQVEEMKT 1551
QY 922 SKLFNFIISLGNKNIYQELIGQSSNFYKILKDSOTFYNESFTNFKSKADDINSL 981
Db 1552 QLELEDELOAAEDAKLRLE-VNMQAMKSOQFERDLQARDEQNEEKRRQLLKOLHEHETEL 1610
QY 982 NDESK-----RKLEEDINKKLTQLSFDLYNK-----YKLL 1015
Db 1611 EDERKORALAAAANKLEVDV-----KLESQVDSANKAREEAKQLRKLOAQMKDYORDL 1666
QY 1016 -----ERLF-----DKKKTGVGRKMQI-----KLTLLKEQLESKLSNLSN 1050
Db 1667 DOAARAREEIFATARENEKAKNLEALQLOEDLAAERARKQADLEKEEAAELASAN 1726
QY 1051 NPKHVLFNSVFNFKKKEAEIAETENTLE-----NTKILKHVYKGLVYNGESSPLKTL 1106
Db 1727 SGRTSLQD-----EKRLEARIAQLEELDEHSNIETMSDRMKAVQQAQQLNNELATER 1782
QY 1107 ESIQOTEDNVASLE--NFKVLKSL-----EGKLDNLNLEKKLSYLSGSLHLLIAELKEVI 1161
Db 1783 ATAQKNENARQOLERONKELRSKLQMEGAVKSF---KSTIAALEAKTASLEQLEQBA 1839
QY 1162 KKNKNTGNSPSENNTDVNNAL---ESYKKFLPEGTVDATVWSBESGDTLEQSQPKKPAST 1218
Db 1840 REKQAAAKTLRQDKKLKDALQVEDERKQAEYKQDA---EKGNLRLKQ---LKROLE 1892
QY 1219 HVGAESENTTITS--QNVDDVEDVDVIVPIEGESEEDYDGLGOVV 1260
Db 1893 EAEEESQRINANRRKLQRELD-----EATESNDALGREV 1926

RESULT 46
EX5B.BORBU STANDARD: PRT: 1169 AA.
ID EX5B.BORBU
AC O51578;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
GN RECB OR BB0633.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=ATCC 35210 / B31;
RA MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Wathley L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi";
RL Nature 390:580-586(1997).
CC -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
```

```
CC ATP) in either 5'- to 3'-or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC -!- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001164; AAC66981.1;
CC HSSP: P56255; 1PJR.
CC TIGR: BB0633;
CC InterPro: IPR004586; RecB.
CC InterPro: IPR000212; UvrD-helicase.
CC Pfam: PF00580; UvrD-helicase; 1.
CC TIGRFAMs: TIGR00609; recB; 1.
CC Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
CC DNA repair; Complete proteome.
CC NP_BIND 18 25 ATP (POTENTIAL).
CC FT SEQUENCE 1169 AA; 137828 MW; B61D63C1C959B91F CRC64;
CC SQ
Query Match 4.0%; Score 338; DB 1; Length 1169;
Best Local Similarity 19.6%; Pred. No. 0.00018;
Matches 288; Conservative 237; Mismatches 455; Indels 490; Gaps 71;
QY 182 NYFDLLRAKLDVNCANDYQCPFNLIKIRANELDVULKLVFGYKRPDLNKNVGMEDY 241
Db 30 NVVINLIKLYSI-----NEILVLTFT---KKATEEMHTRILKVIEN 69
QY 242 IKKNKKTNIENELIEESK---TIDK-----NKNATKEEKKK----- 277
Db 70 AYSNSKTNEILKEAVEQSKKLFISTINFALHNLNFQIETENYSKYKPEKFSKEIDEI 129
QY 278 -----LYQA-----QYDLSYVKNQLEEAHNLISVLEKRIIDTLKKNENIKELDKIN 323
Db 130 VYDFLRKSDSLIQALDIKDYELKFKSDAKTEEIVLKIKKAVE---RDTQELGDWLK 185
QY 324 ETKNPPANGNTPNTLLDKNKKIEEHEKEIKEIAKTIKFNDSLF-----TDPLEEY 378
Db 186 -----TQTAFENILLKKEELIKYKNIIEEDLRMTDEILSFYKNIHQTKGLEIYS 237
QY 379 LRE-----KKNKIDISAKVETKESTEPNEYPNGVTPLSYDINNALNELNSFGD 428
Db 238 KENDIFKIAETLLKKNKFFSTLIEKEFKKSK-----LSPKEL-KIKNDLILGI 285
QY 429 LINFDYTKEPSKNIYTDNERK-----KFI-NEIKEIK-----IEKKIESDKK 472
Db 286 NIKHEKYKSEDRNKNRNLKQVILKVEYKILKYIEKELKKTGTNTIDQNYIISNLK 345
QY 473 SY---EDRSKSLNDITPEYKELLNLEYDSKFNNDNLTNPEKMMGR-----YS 518
Db 346 NYLKSDEK-KLNLATKNRYKIIILIDEAQDLSLIQIEFIIKILTAGIKLFIADPKQIIYS 404
QY 519 YKVEKLTHTHTFASYSKNSKHLEKLTALKYMEDYSLRNIVVEKELKYKKNLSKIENEI 578
Db 405 FRKADISFN-----KEIKNKINTDAR-IVLKINHRSSKKLIGLPLNKIF 447
QY 579 ETLVENIKKDEQLPEFKITKDNKPEKILEVSDIVKVOQVLLMKNKIDELKKTQLIL 638
Db 448 NNIYNNAIADETE---KIDFTNSLPNQK---NDNKNKIVNG-----QETEGINIIT 492
QY 639 KIVLEKHNTHVNSYKQENKQEPYIILVILKKEIDKLKLVMPKVESLINEKKNIKETGOS 698
Db 493 TQTESEEDIY-----QKTALTIKYLAYGKTAENNKI-----RNKMQDTIK 533
QY 699 DNSEPSTEITGQATTKPGQAGSALGDSVQAQEQKQAOQPPVPVPVPEAKAQPPTP 758
Db 534 VILCRGKNEINLIDK-----ALKKEQIQTNKTQEKFL----- 564
```

```
Qy 759 PAPVNNKNTYNSKLDY-----LEKLYEFLNTSYICHKYILVSHSTMNKILKOYKITREE 814
Db 565 -----KTKESEIFYIKCLDRQSPKTLNVLSSKILNPNWLNQRLIKQDKICLIEE 618
Qy 815 SKLSCDPLDLFLFNQNNIPWYSWFDSLNNLSOLFMEIYEKEMVCHLYLKNDKDI-- 872
Db 619 F-----IENIVLEKNEITLINAINKI---TFEKNLWIKIANITDKQKIE 662
Qy 873 -----KNLEEAQKYSTSVKTLN-----SSSQPL-----SLTPQDKPVS 908
Db 663 WAKNKINYKGLLIKEGLENKTYETILEIISKYHKEQNTQSLSLSTLESILNEEPE-- 720
Qy 909 ANDTSHSTNNLSKLFLNLSLGGKNKIYOELIGQKSENFYEKILKDSDTFYNESF- 967
Db 721 --EIEKINNNDNESTIE-LMTIHKSKGLGNIV-----FLNLTPTIENSFF 766
Qy 968 ---TFVYKSKADDINLNDSESKRKLLEDINKLAKTTLQSLFDLYNKYKILERLFDKXK- 1023
Db 767 SKNQFYKYQDG-----KIEYDFKLEE-----NKKYARUKILSEKNI 806
Qy 1024 ---TVGKYKQWIKKLTLLKEQL--ESKLSLNNPKHVLQNFVSF-----FNKKKEA 1069
Db 807 FVVGATRAKFAFIKINSITSKLEIAKIFTIDIKH---DFNIHEIGQKRNKKK-- 861
Qy 1070 EIAETENTLNTKILLKHYKGLVYKNGESSPLKTLSEESTQTDENVASLENFVLSKLE 1129
Db 862 ---YNTNVNTKLT-----PPKPIKNMFKKE--VTS----- 887
Qy 1130 GKLXDLNLLEKKLSYLSLGLHLLIAELKEVKNKNTG--NSPSNTDNNNALESYKFF 1198
Db 888 -----SFSSUTAQAHH-----KEFYENDFKNIYKEKETELDYEPGLE---ET 927
Qy 1189 LPEGTDVATVVSSEGSDFLEQSPKPASTHVGAESENTITTSQNVDDDEVDDVIVPIPF 1248
Db 928 LPKGDIGNIL-----HAAME-----EI-----IFST 949
Qy 1249 SEEDYDDLGQVVTGEAVTPSVIDNLSKIENEYEVLYLKPLAGVYRSKLKLENNVMTFN 1308
Db 950 AKDTFDFN-----KNNIEIEIKQIOLNSLNTIEIQ-----NSLAKMI-YNLTYN 996
Qy 1309 VNVKDLNRSNKRNFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKE-KRDKFLSSVNY 1367
Db 997 IRA---INTRCLDIEELQKEME-----FLIK-----INPEFOKOKYL----F 1031
Qy 1368 IKDSTDTDFNANDVLGYIK-ILSEKYKSLDLSIKKYINDKQGENEKYLPPLNNIETLYK 1426
Db 1032 DKHEDLHKLSD--GYLKGIVDLIFKA---NNKIYILD-----YK 1067
Qy 1427 T---VNDKIDLVHLEAKVLYNTEKSNVEVKIKELNLYLKTQDLADEKKNNNFVGIA 1483
Db 1068 TNYLGNKEDYNTNLENTIKKEYVD---LQYKIYALGIKKILFNKKEY--NOKFGII 1122
Qy 1484 DLSTDYNNHN---LLTKFLSTGWFENLAK 1510
Db 1123 YLFTRAPEDNIECLKSKF-ENGIYF-NLPK 1150

RESULT 47
MYH6_RAT
ID MYH6_RAT STANDARD; PRT; 1938 AA.
AC P02563; Q63351;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MYHC-alpha).
GN MYH6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=Heart;
RX MEDLINE=90016822; PubMed=2798111;
RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT "Complete nucleotide sequence of full length cDNA for rat alpha
RT cardiac myosin heavy chain."
RL Nucleic Acids Res. 17:7527-7528(1989).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=90133919; PubMed=2614840;
RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
RT Comparisons suggest a molecular basis for functional differences."
RL J. Mol. Biol. 210:665-671(1989).
RN [3]
RP SEQUENCE OF 1-167 FROM N.A.
RX MEDLINE=84194059; PubMed=6585819;
RA Mahdavi V., Chambers A.P., Nadal-Ginard B.;
RT "Cardiac alpha- and beta-myosin heavy chain genes are organized in
RT tandem."
RL Proc. Natl. Acad. Sci. U.S.A. 81:2626-2630(1984).
RN [4]
RP SEQUENCE OF 1512-1938 FROM N.A.
RX MEDLINE=82220036; PubMed=7045682;
RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
RT "Molecular characterization of two myosin heavy chain genes expressed
RT in the adult heart."
RL Nature 297:659-664(1982).
RN [5]
RP SEQUENCE OF 1872-1938 FROM N.A.
RC STRAIN=Wistar; TISSUE=Heart;
RX MEDLINE=85179510; PubMed=6241892;
RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
RT "Cardiac myosin heavy chain isozymic transitions during development
RT and under pathological conditions are regulated at the level of mRNA
RT availability."
RL Eur. Heart J. 5:181-191(1984).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15938; CAA34064.1; -
DR EMBL; K01464; AAA41648.1; -
DR EMBL; J00751; AAA41653.1; -
DR EMBL; M32697; AAA41658.1; -
DR PIR; S06005; S06005.
DR PIR; A02988; A02988.
DR HSP; P08799; IWND.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
```

DR InterPro; IPR001609; myosin_head.
DR pfam; PF00063; myosin_head; 1.
DR pfam; PF00612; IQ; 2.
DR pfam; PF01576; Myosin_tail; 1.
DR pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 781 MYOSIN HEAD-LIKE
FT DOMAIN 782 811 IQ.
FT DOMAIN 842 1938 COILED COIL (POTENTIAL).
FT NP_BIND 177 184 ATP.
FT DOMAIN 656 678 ACTIN-BINDING.
FT DOMAIN 758 772 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 789 806 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 815 832 METHYLATION (TRP) (POTENTIAL).
FT MOD_RES 128 128 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 696 696 ALKYLATION (SH-2) (POTENTIAL).
FT MOD_RES 706 706 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 13 13 R -> AP (IN REF. 3).
FT CONFLICT 46 46 V -> A (IN REF. 3).
FT CONFLICT 51 52 VS -> AP (IN REF. 3).
FT CONFLICT 87 87 E -> Q (IN REF. 3).
FT CONFLICT 109 109 MISSING (IN REF. 3).
FT CONFLICT 1566 1566 F -> FF (IN REF. 4).
FT CONFLICT 1575 1575 R -> S (IN REF. 4).
FT CONFLICT 1721 1721 N -> T (IN REF. 4).
FT CONFLICT 1852 1852 T -> N (IN REF. 4).
FT CONFLICT 1870 1870 D -> N (IN REF. 4).
FT CONFLICT 1934 1934 M -> I (IN REF. 4 AND 5).
FT SEQUENCE 1938 AA; 233507 MM; D7BD33FC2B19E3C2 CRC64;
Query Match 4.0%; Score 338; DB 1; Length 1938;
Best Local Similarity 19.6%; Pred. No. 0.00031;
Matches 302; Conservative 262; Mismatches 550; Indels 430; Gaps 71;
QY 28 LVKKLEALEDVLTGYSFOKEKVLNCTSGTAVTTSTPGSKGVASGGSGVASGGS 87
DB 594 LEKNKDPUNETVV---GLYKSSLLK-----MATLFST-----YASADTGDGSGKGG 638
QY 88 VASGSGVASGGSGVASGGSG---NSRRTPNSDSSDAKSVADLKHVRNLYLLTIKELK 143
DB 639 KKKSGFQTVSALHRENKMLNLTHTP-----HFVR--CIIPNERK 680
QY 144 YPOLFD--LTNHLTLCDNIHGFKYLDIGYEINELLY----- 179
DB 681 APGVMDNPLVMHQLRCNGVLEIGIRCKGFP--NRILYGFQRQRYRLNPAAIPEGQFID 738
QY 180 -----KLNFYEDLLRAKLANDVCANDYCOIPN-----L 207
DB 739 SKGAELKGLSLDI-----DHNQYKFGHTKVFVKAGLLGLEEMRDERLSRIIT 787
QY 208 KTRANELDVLKLVGYRKPLD-----NIKDNVG-----KMEDVIK--K 244
DB 788 RQQAARGOLMIEF--KKWVRRDALLYIQWNIAPFGVKWPKWPKLYFKIKPLKSAE 845
QY 245 NKKTIENNELIEESKTTIDKKNKATKEBK--KLYQAQYDLSYNNKQLEEAHNLISYL 302
DB 846 TEKEMANKEEFGRYKDALEKSEARKELEKWSLQEKNDLQ---QVQAEQNDLADA 902
QY 303 EKRIIDLTKNE-----NKKELDKNETKNPPANSNGTPTLLDKNKKIEHEKEIKEL 357
DB 903 EERCQDLINKIQLEAKVKEMTERLEDEE---MNA-----ELTAKRKLIED---ECSBL 951
QY 358 AKTIKFNIDSLTDPLELYLREKNKNIDISAKVE--TKESTEPNPGVTYPLSYNDI 416
DB 952 KKDI-----DDELTFL-----AKVEKEKHATE-----NKV 976

QY 417 NNALNELNSFGDLINPFDYTKEPSKNIYDNERKKFIINEIKKIKIEKKKIESDKKSYED 476
DB 977 KNLTEEMAGLDEIIAKL--TKE-----KKALQEAHQ--ALDDLQAEDEK----- 1017
QY 477 RSKSLNDITKEYEKLLNEIYDSKFNNDIDLTNFEKMMKGRYSKYVE---KLTHHTNTFASY 533
DB 1018 ---VNTLTISKVKLEQQVDD--LEGSLEQKKKVMRDLERAKRKLGGDLKLT--QESIMDL 1070
QY 534 ENSKHNLKLTALKY-----MEDYSLRNIVVEKELKYKKNLISITENIEIETLVNI 585
DB 1071 ENDKLQLEKLLKKFEFDSQKSKIEDQALALQLOKKLKENQAARIEELEELE--AERT 1128
QY 586 KKEDEQLFEKKTIDENKPDKEKILEVSDIVKQVKVLLMKN--IDELKKTOLILKNVEL 643
DB 1129 ARAKVEKLRSOLTRELEISERLEAGGATSVQIE---MNKKREAEQKMRDRLEEAFL 1184
QY 644 KNIHVPNSYKOENKQEPYLLIVLAKKEDKLKVFMPKPVESLINEEKNKIKTEGQSDNSP 703
DB 1185 QHEA---TAAALRRKKHADSVAELEGQIDNLQ---RVKQLEKSEKSEKLE--LDDVTS 1234
QY 704 STEGEITGOATTPGQQAAGSALLEGDSVQAQAEQKQAQPPVPVPVPEAKAQV---PTPPA 760
DB 1235 HMEQIIKAKANLEKYSR-----TLEQANEYR-----VKLEAQRSLNDDFTQRA 1279
QY 761 PYNKNTEN---VSKLDYLEKLEFNTSYCHKYILVSHSTMNEKILKQYKIKTEESKL 817
DB 1280 KL--QTENGELARQLEKEALIIWLTRGKL-----SYTQOMEDLKROL-----BEEGRA 1326
QY 818 SS-----CDPLDLFNIQNNIPVMYSMPFDSLNNLSQLFMEIYKEMVCLYK 865
DB 1327 KNALAHALQSAHRHDCDLUREQYEEEMEAKAEQLQSVLSKANSEVAQWRKYETDAIORTEE 1386
QY 866 LKD-NDKIKNLLEEAKKVSTSVKTLSSSSMQPLSTLPDQKPE---VSANDDTSHSTNLNN 921
DB 1387 LEEAKKLAQRLQDAEAEVAVNA-KCSLEKTKHRLQNEIEDLMVDVRSNAAALDK 1445
QY 922 SUKLFENILSGKNKIYQELIGQKSS-----ENYEKILKSDTFYNES 966
DB 1446 KORNFDKILAEWKQK--YBESQSELESQKARSSTELFKLNAYEESLEHLETFKREN 1503
QY 967 FTNFVKSKADDINSIND-----ESKRKKLEEDINKLKKTL----- 1001
DB 1504 -----KNQOEESLDTQELGEGGNVHELEKIRKQLEVEKLELQSALEAEASLEHEEGK 1558
QY 1002 -----QLSFDLYNKYKLER-LDFDKKTVGYKYM-OIKKLTLLKBOLESKLSLNNPKHV 1055
DB 1559 ILRAQLEF---NQIKAEIERKLAEKDEEMEQAKRNHLRVVDSLOTSLDAETRSRNEALRV 1615
QY 1056 LQNFVFFNKKEAEIAETENTLNTKILL---KHVGLVKYNYGCESSPLKTLSEESIQ 1111
DB 1616 -----KKMEGDLNEMEIQLSQANRIASBAQKHLK-----NAQAH-----LKDTQLQ 1657
QY 1112 TEDNYASLENPKVLSKLEGLKADNLNLEKKLSYSSGLHLLHIAELKEIKVKNKNTGNSP 1171
DB 1658 LDDAVRANDD-----LKENAIVERRNTLQAELEELRAVVEQTERSKLAEQEL 1707
QY 1172 SENNTDV-----NNALESYKFL-PEGTDVATVVSSEGSTLEQSPKPKPASTHVGAES 1224
DB 1708 IETSERVOLLHSQNTSLINQKKMDADLSQLQTEVEEAVQECRNEAEKAKKAITDAAMMA 1767
QY 1225 NTITTSQ-----NVDDEVDDVIIVITFGSEEDYDDLQGV-VTGEAVTSPVIIN 1272
DB 1768 BELKKEQDTSALHERMKNMEQTIKDL-----QHRLEDAEQIALKGGKQLQKLEA 1818
QY 1273 ILSKTENEYEVLYLPLACVYRSLKOLENNYMTNVNVKIDILNSPKNRENPKNVLES 1332
DB 1819 RYRELENELE-----AEQKRNAESVKGMRKSKERRIKELTYQTEEDKN----- 1861
QY 1333 LPYKDLTSSNYVVKDPYKFLNKEKRDKF---LSSVNYIKDSID 1373
DB 1862 LVRLQDLVDKQLVKYAKRQAEAEAEQANTNLSKFRKVQHELD 1905

FT	CONFLICT	1998	1998	G -> R (IN REF. 5).
SQ	SEQUENCE	2704	AA; 313084	MW; A7219B687A634A77 CRC64;
	Query Match	4.0%;	Score 338;	DB 1; Length 2704;
	Best Local Similarity	19.7%;	Pred. No. 0.00045;	
	Matches	292;	Conservative	233; Mismatches 514; Indels 448; Gaps 63;
Qy	166	YLIDGYE-----EINELLYKLN-----FYDILL-----RAKNDVCANDYQCOIFENLKIR	210	
Db	715	FLEDSQESQVSGFSDDITQLEKEVAVCKQYQYQELLSAEREQESVYNYLISEVRNIRI	774	
Qy	211	AN-----ELDVLLKLVF-----GYRKPLDNIKDNVG-----KMEDYIKNNK	246	
Db	775	LENCEDLRIQIRTPLERDDLHESVFRITEQEKLLKELERLKDGLTTNKCEFFSAA	834	
Qy	247	KT-----ININELIESKKTIDKNNATKEEKKKLYQAOYDLSYNNKOLE	293	
Db	835	ASSSVPTLRSELNVLOMNOVYSSSTYIDKYVTA-----ISIV-----	874	
Qy	294	EAHNLISVLE-KRDTLTKNENIKELLDKINEIK-----NPPANGNTPN--TLLDNKN	345	
Db	875	KGYCLISFLRLKTVNLVLTQAAEALVKLYETKLCEEEAVIADKNNIENLISLTKQWRS	934	
Qy	346	KIEBH-----EKEIEIAIKTIKENIDSLFTD-----PLELEY-----LREKNNKNDI	388	
Db	935	EVDEKROVFALEBLOK-AKAIS---DEMFKTYKERDLDFDWHKEADQLVERQNVHV	990	
Qy	389	SAKVETESTEPNEYPNGVTPPLSYNDINNALNELNSFG-----DLINFPD-----	434	
Db	991	-----YTKEPSKNIYTD-NERKKFTNEIKIEKIKIEKKIESDKKSYEDRSKSLNDI-	1024	
Qy	435	-----TTORKIOENQENSKTATQLNQOKMLVSELEMK---QSKMDECOKVAEOYSATVKDYE	1080	
Qy	485	-TKYEKLLMNEIYDS-----KFNNIDLTNFEKM-MGKRYSYKVEKLTHTINTFAS-----	532	
Db	1081	LQMTYRAMVDQOQSPVKRRRMOSSADLLIQEFMDLRTYRTALVTLMTOYIKFAGDSLK	1140	
Qy	533	-----YNSK-----HNLEK	542	
Db	1141	RLREEVVIINSEHPVNIKELEIKRCKETSEHGAYSDLLQROKATVLENSKLTGKISELER	1200	
Qy	543	LTKALK-----YMEDYSLRNIVVEKEKLYKYNLISKI--	574	
Db	1201	MVAELKKQKSVEBELPKVREAENELRKQORNVEDI SLQIKRAESEAQYRRELETIVR	1260	
Qy	575	-----ENIE-----TLVENIKKD--EQLFEKKTIDKENPDEKILEVSDIVKVQOVKV	622	
Db	1261	EKEAARELERVROLTTEAEAKRAAVEENLLNFRLQLEENTFTTRTLE-----	1308	
Qy	623	LLMNKIDELKKTQLILKNVE-LKNIHVPNSYKQENKQEPYVILVKKIEIDKLKVMFKV	681	
Db	1309	-----DHLKRRKOLSLNLEQOKNKLMEELRRKRDNEE--LLKLKIQMEKDLAFOKV	1359	
Qy	682	ESLNEEKKNKTTGQSDNSPSTEGEITGOATYKPG---QOASALBEGDSVQAOAQOQK	738	
Db	1360	AEOLKEKQKTELEAR---RKITEIQYTCRENALPVCPIQTQASCRAVTTGLQOEHDKQK	1415	
Qy	739	QAQPPVPVPEAKAOVPTPPAPVNNKTENVSKLDYLEKLYEFNTSYICHYILVSHST	798	
Db	1416	-----AEELKQVDELTA-ANRKAQO-----DMRELTVE-LNA-----	1446	
Qy	799	MNEKILKOYKTKTBEEKSKLSCDPLDLFFNTQNNIPVMYSFDSLNNLSQLFWEIYBKE	858	
Db	1447	-----LOLEKTSSEKARL-----LKDKLDNETNTLRCLKLELRKD	1483	
Qy	859	MVCNLYKLLKONDKIKNI-----LEEAKVSTSVKTLSSSSWQPLSLTPQDKPEVSANDD	912	
Db	1484	QAEQYQSOQLRELGRQLNQTGKAEAAQEAQEAQASDLKKIKRNTOLESLNHEKGLQREVD	1543	
Qy	913	---TSHSTNLNLSKLFIENILSLGKNKIYQELIGQKSSSENFYEKILKDSOTFYNESFTN	969	

Db 1544 RITRAHVAEKNIQHLSQHSFDEKELERLQICQRCSOHLRKEOFEPS-----HEQLLQ 1598

Qy 970 FVKSKA---DDINSLNDSKRKKLEEDINDLKKTQLUSFDLYNNKYKLLKLRLFPDKKTVG 1026
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 1599 NIKAEKENNDKIQLNEE--LEKSNECAEMLKQVEBELTRQNNETFKLMORIQAESNIV 1656

Qy 1027 KYKWQIHK---LTLLEQLESKLNSLNPNKHVLQNFVSFFNK--KEAIEAETENTLE 1079
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 1657 LEKOTIOORCALQIADGFKDQURSTNEHLHKTKTEQDFQRKICLEEDLAQSQNLSV 1716

Qy 1080 NTK-----ILLKHGYGLVYNGE---SPLKTLSEESIQTEDNYASLENFKVLSKLE 1129
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 1717 EFQKCDQONIIIONTKKEVRNLNAELNASKEEKRRGEQKVQLQAQVOELNNR-LKKVQ 1775

Qy 1130 GKL-----KDNLEKKKLSVSSGLHLTAE-----LKEVIKNKNTGNSPSENNTDVN 1179
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 1776 DELHLKTEEQMTHRKVMVLFQEESGPKQSAEFPRKKMEKLMESKITENDISGIRLDF- 1834

Qy 1180 NALESYKKFLPEGHVDVATVVSSESGDTLEOSQPKKPASTHVGAESNTITTSNQNVDDVEDD 1239
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 1835 VSLQENSRAOENAKLCETNIKELEROLQQVREQMOQGQM--EANYHOKCKLEDE--- 1889

Qy 1240 VIIPIFGESEDDDLGOVVTVGEAVTPSVIDNLSKI-----ENEYEVLKPLAGVYR 1294
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 1890 -----LIAQKRE-----VENLQKMDOQOIKGHEHQLVLLQ-----C 1920

Qy 1295 SLKQLENNVMT---FNVNVKDILNRFNKRNFKNVLESDDLTPYKDLTSSNVVWKDPY 1350
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 1921 EIQKSKTAKCTFPDDEMVTKECHGSELSSRTGHLHTPRSPFLLRWTQEPQPLEKW 1980

Qy 1351 KF-----LNKEKRDKFLSYNYIKDSIDTDINF 1378
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 1981 QHRVVEQIPKEVFQPPGAPLEKEKSQCXYSEY-FSQTSTELQITF 2025

RESULT 49

Y373_HUMAN ID Y373_HUMAN STANDARD; PRT; 1539 AA.

AC O15078;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein KIAA0373.

GN KTA00373.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=37349984; PubMed=9205841;

RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
Miyaajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";

FT DNA Res. 4:141-150(1997).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AB002371; BAA20828.1; "

KW Hypothetical protein; Coiled coil.

FT DOMAIN 18_1514 COILED COIL (POTENTIAL).

SQ SEQUENCE 1539 AA; 180065 MW; D901314E961BF001 CRC64;

Query Match 4.0%; Score 337.5; DB 1; Length 1539;

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X64838; CAA46050.1; -
CC DR EMBL: M97501; AAA35693.1; -
CC DR PIR: S22695; S22695.
CC DR Genew: HGNC:10461; RSN.
CC DR MIM: 179838; -
CC DR InterPro: IPR000938; CAP-Gly.
CC DR InterPro: IPR001878; Znf.CCHC.
CC DR Pfam: PF01302; CAP_GLY_2.
CC DR SMART: SM00343; Znf.C2HC; 1.
CC DR PROSITE: PS00845; CAP_GLY_1; 2.
CC DR PROSITE: PS0245; CAP_GLY_2; 2.
CC KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
CC FT DOMAIN 78 120 CAP-GLY 1.
CC FT DOMAIN 143 204 SER-RICH.
CC FT DOMAIN 232 274 CAP-GLY 2.
CC FT DOMAIN 304 331 SER-RICH.
CC FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
CC FT DOMAIN 1408 1421 CCHC-BOX.
CC FT VARSPLIC 457 491 MISSING (IN SHORT ISOFORM).
CC FT CONFLICT 1069 1069 D -> E (IN REF. 2).
CC SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;

Query Match 4.0%; Score 336.5; DB:1; Length 1427;
Best Local Similarity 19.3%; Pred. No. 0.00025;
Matches 283; Conservative 249; Mismatches 508; Indels 425; Gaps 60;

QY 43 YSLFQKEKVLNDCSTGTAVTTSFG-SKGSVASGGSGSVASGSGSVASGSGSVA 101
DB 269 YGLFAPVHVKTIGFPST-----TPAKANAVRRVMATTSASLKRSPASSLSMSVA 323
QY 102 SGGSGSRRTNPDSNSSSDAKSYAD-----LKHVRN--YLLTIKELKYPQLFDLTN 152
DB 324 SSVSSRPSRTGLLTETSSRYARKISGTTALQEALEKQHQHQLAERDLERAFAKATS 383
QY 153 HM-----LTLCDNIHGFYLDGVEE-INELLYKLNFYDILLRAKLVNVCANDYQCIPF 205
DB 384 HVGEIEQELALAR-----DGHQHVLELEAKM---DQLRTMVE---AADREKVEL 427
QY 206 NLKIRANELDVLKLVFGYRKPLDNIDKONVGMEDYIKNKKTNIENINELIESKKTIDK 265
DB 428 -----LNQLEEBKRVVEDLQFVBEESTYKGLDETQTKLEHARIKELOSLLFEKTKADK 482
QY 266 NKNATKE-----EKKKLYQAYDLISYNKOLEAHNLISVLEKRIDPLKKNENIKELL 319
DB 483 LQRELEDTRVATVSEKSRIMELEKDLALRVQEAEE-----LRRLESNKPGADVMSL 535
QY 320 DKINEIKNPPANGNTPNTLLDKNKKIE-----EHEKEIKEIAKTIKFNIDSLFTDPLEL 375
DB 536 SLQLEISS-----LQKLEVTTRTDHQREITS-----LKE 564
QY 376 EYLLREKKNIDISA-KVETKESTPEYNGVTPYPLSYNDINNALNELNSFGDLINPFD 434
DB 565 HFGARETHQKEIKALYTATEKLSKENE-----SLKSKLEHAN----- 602
QY 435 YTKEPSKNYITDNERKFFINEIKIEKIKIESDKKSYEDR-----SKSLNDITPK 486
DB 603 -----KENSVDIALWKSLETAISHQOAMEELKVSFSKGLGTETA 643
QY 487 EYEKLLNEYDYSKFNNDLTNFEKMMGKRYSKYKVEKLTHTHTFASYSKHNLEKLTKA 546

DB 644 EFAELKTOI-----EKMR-LDYQHEIENLQNOO-----DSERAAHAKEMEALR 685
QY 547 LKYMEDYSLRNIVWELKYKKNLISKIENETETIVENIKKDEQ-LPEKKTITKDNKPD 605
DB 686 AKLMK-----VIKEK-----ENLEATRSKLDKAEDOHVEMEDTL--NKLQ 725
QY 606 EKILEVSDI-----VKVQVQVLLMKNKIDELKKTOLIKNLVELKHNHVPNSYKQENKQBP 661
DB 726 EAEIKVKEVLVLAQKNEQTKVIDNFTSQLKATEKEKLLDLD----- 766
QY 662 YLIVLKEIDKLKVPMPKVESLINEEKNAIKTEGOSDNSEFSTEGEITGQATTPGQQA 721
DB 767 ----ALRKASSGCKSEMCKLROOLEAAEKQIKHLETEKNAESKASITR----- 812
QY 722 GSALGSDSVQAQAEQKQAQPPVPVPEAKAQQVTPPPAPVNNKTNVSKL-----DYLEK 777
DB 813 -----ELQRELK-----LTNLOENLSEVSOVKRETLK 840
QY 778 LYEFNTSYICHYILVS-HSTMNEKILKQYKITKEESKLSGCDPLDLFIQNNIPVM 836
DB 841 ELQILKEFAEASEEAVSVORSQETVKNLHQ--KEEQFNMLSSD----LEKLENLADM 894
QY 837 YSMFDLNNLSQLPMEIYEKEMVCLNYKLDKNDXIKNLEEAQKVVSTSVKTLSSSSMOP 896
DB 895 EAKFREKDEREEQL-----IKAKELENDIAEIMKMS----- 926
QY 897 LSLTPQDRPEVSANDDTSHSTNNLSKLFIENILSGKNKNIYO-ELIGOKSEN--EYE 953
DB 927 -----GDNSSQLTKMDELRL-----KERDVEELQOLKUTKANENASFLQ 965
QY 954 KILKSDTFYNESFTNFVKSKADDINSLNDESKRKKLEEDINKKKTLOLSFDLNYKYL 1013
DB 966 KSIDETVKAESQOEAAK-----KHBEKKELEKRLSDLEKKMETSHNOQELKA 1016
QY 1014 KLERLFDKKKTVGYKMQIKKLTLLKEQLESKLNLPKH-VLQNFVFFPKKKEAETA 1072
DB 1017 RYERATSETKTKEHILQNLQTL--DTEDEKLKGARENSGLLOELE---ELRKQADKA 1071
QY 1073 ETEVNTLENTKILLKHYKGLVKGNGESSPLKTLSEESIOTEDNYASLENFKVLKLEGL 1132
DB 1072 KAAQTAEDAMQIMEQ-----WTKEKTEF---LASLEDTK---QTNAKL 1108
QY 1133 KDNLEKKKLSYLSGLHLHIAELKEVIKNKNYTGNSPENNTDYNNALESYKFKPLPEG 1192
DB 1109 QNELDTLKE-----NNLKNVEELNKS-----ELLTVENOKMEEFRKEIETL 1150
QY 1193 TDVATVWSESGSDTLEQSQPKPASTHVCAESNTITTSQNVDE---VDDVII-----V 1243
DB 1151 KOAAQAKSQOLS-ALQOENVK--LABELGRSDEVTSHOKLEERSVLNQLEMKKRES 1207
QY 1244 PIFGESEEDYDGLGVVTGEAVTPSVIDNLSKIENEYEVL-----YLKPLAGVYRSL-- 1296
DB 1208 KFIKDADEKASLQKSISITSALLTEKDAELEKLRNEVTVLRGENASAKLSHVQVTL 1267
QY 1297 -KKOLENNVMTNVAVKOTILNSRFNKRNFKNVLESDLIPYKDLTSSNVVVKDPYKFLNK 1355
DB 1268 DKVKLELKVKNLELQKE-----NKRQLSSSGNTDTQADEDERA-----Q 1308
QY 1356 EKRDPLSLSYNIKOSIDTIDINFAN-DVLGYKYLISE-KYKSDLSIKKIYNDKQGENEK 1413
DB 1309 ESQIDFL-----NSVIVDLQKQDKLMMKVMEMSEALNGDGLNNYSDDOEKQSK 1361
QY 1414 YPFLNNIETLYKTVNDKIDLEVIH 1438
DB 1362 KKP-----RLFCIDICDFDLH 1377

RESULT 51
DMD_HUMAN STANDARD; PRT; 3685 AA.
AC P11532; Q14169; Q14170;
DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN Dystrophin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88194521; PubMed=3282674;
RA Koenig M., Monaco A.P., Kunkel L.M.;
RT "The complete sequence of dystrophin predicts a rod-shaped
cytoskeletal protein.";
RL Cell 53:219-228(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89345106; PubMed=2668885;
RA Rosenthal A., Speer A., Billowitz H., Cross G.S., Forrest S.N.,
RA Davies K.E.;
RT "Two human cDNA molecules coding for the Duchenne muscular dystrophy
(DMD) locus are highly homologous.";
RL Nucleic Acids Res. 17:5391-5391(1989).
RN [3]
RP SEQUENCE OF 1-497 FROM N.A.
RX MEDLINE=87273512; PubMed=3607877;
RA Koenig M., Hoffman E.P., Bertelson C.J., Monaco A.P., Feener C.,
RA Kunkel L.M.;
RT "Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and
preliminary genomic organization of the DMD gene in normal and
affected individuals.";
RL Cell 50:509-517(1987).
RN [4]
RP SEQUENCE OF 404-1137 FROM N.A.
RX MEDLINE=88111512; PubMed=3428261;
RA Cross G.S., Speer A., Rosenthal A., Forrest S.M., Smith T.J.,
RA Edwards Y., Flint T., Hill D., Davies K.E.;
RT "Deletions of fetal and adult muscle cDNA in Duchenne and Becker
muscular dystrophy patients.";
RL EMBO J. 6:3277-3283(1987).
RN [5]
RP SEQUENCE OF 665-722; 2098-2204 AND 2305-2366 FROM N.A.
RX MEDLINE=89083552; PubMed=3205741;
RA Chamberlain J.S., Gibbs R.A., Ranier J.A., Nguyen P.N.,
RA Caskey C.T.;
RT "Deletion screening of the Duchenne muscular dystrophy locus via
multiplex DNA amplification.";
RL Nucleic Acids Res. 16:11141-11156(1988).
RN [6]
RP SEQUENCE OF 2147-2204 FROM N.A.
RX MEDLINE=89345155; PubMed=2569720;
RA Blondin L.A.J., den Dunnen J.T., van Paassen H.M.B.,
RA Wapenaar M.C., Grootscholten P.M., Ginjaar H.B., Bakker E.,
RA Pearson P.L., van Ommen G.J.B.;
RT "High resolution deletion breakpoint mapping in the DMD gene by whole
cosmid hybridization.";
RL Nucleic Acids Res. 17:5611-5621(1989).
RN [7]
RP SEQUENCE OF 2305-2364 FROM N.A.
RX Huth A., Will K., Speer A., Bauer D.;
RA Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [8]
RP REVIEW ON DMD POINT MUTATION VARIANTS.
RX MEDLINE=95038763; PubMed=7951253;
RA Roberts R.G., Gardner R.J., Bobrow M.;
RT "Searching for the 1 in 2,400,000: a review of dystrophin gene point
mutations.";
RL Hum. Mutat. 4:1-11(1994).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=94320940; PubMed=8045556;
RA Rininsland F., Reiss J.;
RT "Microlesions and polymorphisms in the Duchenne/Becker muscular
dystrophy gene.";
RN Hum. Genet. 94:111-116(1994).
RN [10]
RP VARIANT DMD ARG-54
RX MEDLINE=94004962; PubMed=8401582;
RA Prior T.W., Papp A.C., Snyder P.J., Burghes A.H.M., Bartolo C.,
RA Sedra M.S., Western L.M., Mendell J.R.;
RT "A missense mutation in the dystrophin gene in a Duchenne muscular
dystrophy patient.";
RL Nat. Genet. 4:357-360(1993).
RN [11]
RP VARIANTS DMD HIS-365; TRP-2191 AND ARG-2937.
RX MEDLINE=95152525; PubMed=7849724;
RA Nigro V., Nigro G., Esposito M.G., Comi L.I., Molinari A.M.,
RA Puca G.A., Politano L.;
RT "Novel small mutations along the DMD/BMD gene associated with
different phenotypes.";
RL Hum. Mol. Genet. 3:1907-1908(1994).
RN [12]
RP ALTERNATIVE SPLICING (DYSTROPHIN-1 AND -2).
RX TISSUE=Retina;
RA White R.A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [13]
RP ALTERNATIVE SPLICING (DYSTROPHIN-3).
RX TISSUE=Brain;
RA Feener C.A., Koenig M., Kunkel L.M.;
RT "Alternative splicing of human dystrophin mRNA generates isoforms at
the carboxy terminus.";
RL Nature 338:509-511(1989).
CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
PLASMA MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS ARE PRODUCED BY
ALTERNATIVE SPLICING; THE 3 SHORTER VARIANTS ARE CALLED
DYSTROPHIN-1, -2 AND -3.
CC -!- DISEASE: Defects in DMD are the cause of Duchenne muscular
dystrophy (DMD) and Becker muscular dystrophy; a sex-linked recessive
most common form of muscular dystrophy; a sex-linked recessive
disorder. It typically presents in boys aged 3 to 7 year as
proximal muscle weakness causing waddling gait, toe-walking,
lordosis, frequent falls, and difficulty in standing up and
climbing up stairs. The pelvic girdle is affected first, then the
shoulder girdle. Progression is steady and most patients are
confined to a wheelchair by age of 10 or 12. Flexion contractures
and scoliosis ultimately occur. About 50% of patients have a lower
IQ than their genetic expectations would suggest. There is no
treatment. BMD resembles DMD in hereditary and clinical features
but is later in onset and more benign.
CC -!- DISEASE: Defects in DMD are a cause of X-linked dilated
cardiomyopathy (XLCM).
CC -!- MISCELLANEOUS: THE DMD GENE IS THE LARGEST KNOWN GENE IN HUMANS.
IT IS 2.4 MILLION BASE-PAIRS IN SIZE AND COMPRISES 79 EXONS.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.
CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC -!- DATABASE: NAME=DMD; NOTE=Dystrophin Mutation Database;
WWW="http://www.dmd.nl/database.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06179; CAA29545.1; ALT_SEQ.
CC EMBL; X06178; CAA29544.1; -.

Db 2427 LAPGLTTIGASTQVTLVTPVTKETAISKLEMPSSMLLEVPALADFNRAWTELTDL 2486
QY 1256 --LGQVTTGEAVTPSIDNLSKIENEVEVL-----YKPLAGVYRSKKKQLENNVM 1305
Db 2487 SLLDQVIKRSQVWVGDLDEINEMIIKQKATQMDLEQRPOLEELITAAQNLKNK----- 2540
QY 1306 TFWNVKQILNRSFKNREKFNKVDLPYKDLTSSNVVVKDPYKFLN-KEKRDKEFLSS 1364
Db 2541 TSNOEARTIITDRIERIQNDQVOEHLQNRQ--QLNEMLKDSTQWLEAKEEAEQVLGQ 2598
QY 1365 YN-----YIKDSI-----DTDINFAND-----VLGYIK----- 1387
Db 2599 ARAKLESWKEGYTVDATQKKTETKQAKLDQRQWQTNVDVANDLAKLLRDYSADDTFK 2658
QY 1388 --ILSEKYSKSDLSIKKYINDKQ-----ENKYLPLNNIETLYKTVDNK 1431
Db 2659 VHMITEINASWRSIHKVRSEREALEETHRLLOQFPIDLDLEKFLAWLTEAETANVLQD- 2717
QY 1432 IDLFVHLEAKVNTYKESNVVEKIKELNYLKTQDKLADPKKNNFVGIADLSTYNNH 1491
Db 2718 -----ATKRELEDSKGVKELM--KQWDLQGEIEAHTDVVH 2753
QY 1492 NNLLTKFLSTGMVFENLAKTVLS-----NLLDGNLQGM-----LNISQH 1530
Db 2754 N-----LDENSOKILRSLEGSDDAVILLQRRLDNNMFKWSLLRKKSLNIRSH 2799

RESULT 52
DESP_HUMAN
ID DESP_HUMAN STANDARD: PRT: 2871 AA.
AC P15924; Q14189; O75993; Q9UHN4;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Desmoplakin (DP) (250/210 kDa paraneoplastic pemphigus antigen).
GN DSP.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM DPI).
RC TISSUE=Foreskin;
RX MEDLINE=921115697; PubMed=1731325;
RA Virata M.L.A., Wagner R.M., Parry D.A.D., Green K.J.;
RT "Molecular structure of the human desmoplakin I and II amino
RT terminus";
RL Proc. Natl. Acad. Sci. U.S.A. 89:544-548(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM DPI).
RC Phillips S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1120-2871 FROM N.A. (ISOFORM DPI).
RC TISSUE=Foreskin;
RX MEDLINE=90153880; PubMed=1689290;
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
RA Angst B.D., Nilles L.A.;
RT "Structure of the human desmoplakins. Implications for function in
RT the desmosomal plaque";
RL J. Biol. Chem. 265:2603-2612(1990).
RN [4]
RP ERRATUM.
RX MEDLINE=90361712; PubMed=2391353;
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
RA Angst B.D., Nilles L.A.;
RL J. Biol. Chem. 265:11406-11407(1990).
RN [5]
RP SEQUENCE OF 2854-2871 FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=20062965; PubMed=10594734;
RA Whitlock N.V., Ashton G.H., Dopping-Hepenstal P.J., Gratian M.J.,
RA Keane F.M., Eady R.A.J., McGrath J.A.;

"Striate palmoplantar keratoderma resulting from desmoplakin
haploinsufficiency";
J. Invest. Dermatol. 113:940-946(1999).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=98012209; PubMed=9348293;
RA Kowalczyk A.P., Bornslaeger E.A., Borgwardt J.E., Palka H.L.,
RA Dhaliwal A.S., Corcoran C.M., Denning M.F., Green K.J.;
RT "The amino-terminal domain of desmoplakin binds to plakoglobin and
RT clusters desmosomal cadherin-plakoglobin complexes.";
RL Cell Biol. 139:773-784(1997).
CC !- FUNCTION: MAJOR HIGH MOLECULAR WEIGHT PROTEIN OF DESMOSOMES.
CC INVOLVED IN THE ORGANIZATION OF THE DESMOSOMAL CADHERIN-
CC PLAKOGLIBIN COMPLEXES INTO DISCRETE PLASMA MEMBRANE DOMAINS AND IN
CC THE ANCHORING OF INTERMEDIATE FILAMENTS TO THE DESMOSOMES.
CC !- SUBUNIT: HOMODIMER.
CC !- SUBCELLULAR LOCATION: INNERMOST PORTION OF THE DESMOSOMAL PLAQUE.
CC !- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DPI/DPI1 (SHOWN HERE) AND
CC DPII/DPI2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC !- TISSUE SPECIFICITY: DPI IS APPARENTLY AN OBLIGATE CONSTITUENT OF
CC ALL DESMOSOMES; DPII RESIDE PREDOMINANTLY IN TISSUES AND CELLS OF
CC STRATIFIED ORIGIN.
CC !- DOMAIN: THE N-TERMINAL REGION IS REQUIRED FOR LOCALIZATION TO THE
CC DESMOSOMAL PLAQUE AND INTERACTS WITH THE N-TERMINAL REGION OF
CC PLAKOPHILIN 1. THE C-TERMINAL REGION INTERACTS WITH INTERMEDIATE
CC FILAMENTS.
CC !- DISEASE: DEFECTS IN DSP ARE A CAUSE OF STRIATE PALMOPLANTAR
CC KERATODERMA II (PPK2, KPP2 OR SPPK2), CHARACTERIZED BY SKIN
CC THICKENING IN THE PALMS (LINEAR PATTERN) AND THE SOLES (ISLAND-
CC LIKE PATTERN) AND FLEXOR ASPECT OF THE FINGERS; AND RARELY BY
CC ABNORMALITIES OF THE NAILS, THE TEETH AND THE HAIR.
CC !- SIMILARITY: CONTAINS 17 PLECTIN REPEATS.
CC !- SIMILARITY: CONTAINS 2 SPPECTRIN REPEATS.
CC !- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M77830; AAA85135.1; -
CC EMBL; AL031058; CAA19927.1; -
CC EMBL; J05211; AAA35766.1; -
CC EMBL; AF139065; AAF19785.1; -
CC PIR; A35536; A35536.
CC Genew; HGNC:3052; DSP.
CC MIM; 125647; -
CC InterPro; IPR001101; Plectin_repeat.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00681; Plectin; 10.
CC SMART; SM00250; PLEC; 17.
CC Repeat; Coiled coil; Phosphorylation; Cytoskeleton;
CC Structural protein; Alternative splicing.
CC DOMAIN 1 1056
CC DOMAIN 1057 1945
CC DOMAIN 1946 2871
CC REPEAT 347 447
CC REPEAT 858 952
CC REPEAT 1018 1945
CC REPEAT 2009 2045
CC REPEAT 2046 2083
CC REPEAT 2084 2121
CC REPEAT 2122 2159
CC REPEAT 2163 2197
CC REPEAT 2198 2233
CC REPEAT 2251 2288
CC REPEAT 2289 2326
CC REPEAT 2327 2364
CC REPEAT 2365 2402

```
FT REPEAT 2406 2440 PLECTIN 11.
FT REPEAT 2456 2493 PLECTIN 12.
FT REPEAT 2507 2544 PLECTIN 13.
FT REPEAT 2610 2647 PLECTIN 14.
FT REPEAT 2648 2685 PLECTIN 15.
FT REPEAT 2724 2761 PLECTIN 16.
FT REPEAT 2762 2799 PLECTIN 17.
FT DOMAIN 2824 2847 6 X 4 AA TANDEM REPEATS OF G-S-R-[SR].
FT DOMAIN 1 584 INTERACTS WITH PLAKOPHILIN 1 AND JUNCTION
FT MOD_RES 2849 2849 PLAKOGLOBIN
FT VARSPIC 1195 1794 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT CONFLICT 905 905 MISSING (IN ISOFORM DP11).
FT CONFLICT 1120 1120 R -> A (IN REF. 2).
FT CONFLICT 2687 2688 SV -> RL (IN REF. 2).
SQ SEQUENCE 2871 AA; 331771 MW; 5639B5B7CD4690B7 CRC64;

Query Match 4.0%; Score 334.5; DB|: Length 2871;
Best Local Similarity 18.4%; Pred. No. 0.00062;
Matches 294; Conservative 276; Mismatches 552; Indels 477; Gaps 64;

QY 105 SGNRSRTNPSDSSDSDAKSYADLKHVRNY-----LLTIKELKYPQLFDLTNHHMLTLC 158
DQ 938 SKRDKSEEVQKTAELCAINSIKDYELQASYTSGLETLLNI-PIKRTMTQSPSGVILQEA 996
QY 159 DNIHGPKYIDGYEINELLYKLFYDILLRAKLDVNCANDYCQIPFNKIRANELDVLK 218
DQ 997 ADVHA-RYI-----ELLTRSGDYREFLEMLKSL-----EDLKLKNTKIEVLE 1038
QY 219 KLVFGYRKPLDNKDNVGMEDYIKKNKTYIENINELIBESKKTIDKNKNATKEEKKKL 278
DQ 1039 EEL---RLARDANSEN-----CNKNKFLDNQLQYQAECSOF--KAKLASLEELKR-- 1084
QY 279 YQAQYDLSTYNKQLEBAHNLISVLEKRIDTL-----KKNENIKELDKINEIKNPPANS 333
DQ 1085 -QAELDGSAKQNDKCYQIKELNEKITRLTYEIDEKRRKRSVEDRFDOQND----- 1138
QY 334 GNPNTLLDNKKKIBEHKEIKIAKTIKFNIDSFTDPLEYLYLRKNKNIDISAKVE 393
DQ 1139 -----YDQLQKAROCEKE-----NLGWQKLESEKAIKEYEIERLRLVLL 1178
QY 394 TKESTEPNYPNGVTVPLS-YN-DINNALN---ELNSFGDLINPDYTKE-PSKNI--- 443
DQ 1179 QEGTKRREYNELAKVRNHYNEENSLRNKYETENITKTIKEISMOKEDDSKNLNRQ 1238
QY 444 -----YTDNERKKFEINKEK-----IKIEKKKI 467
DQ 1239 LDRLSRENRLKDEIVRLNDSILQATEQRRRAEENALQKACGSEIMQKQHLEIELQV 1298
QY 468 ESD-----KKSIEDRSKSLNDITKEYEKLLEIYD-----SKFNNDILT 507
DQ 1299 MOORSEDNARHKOSLEAAKTIQDKNKETERLKAETFEAEKARWEYENELSKVRNNDY-- 1356
QY 508 NFEKMMGKRYSKVE---KLH-HNTPASYENS---KHNLKLTAKALYMEDYSLRLNIV 559
DQ 1357 ---EEIISLKNQFETENITKTIHQLTMQKEEDTSYRAQIDNLTR-----ENRS 1404
QY 560 VEKELYYKNLISKIENETELVENIKKDEQLEFKKTIKDKENKDPDEKILEYSDIVKQV 619
DQ 1405 LSEIKRLKNTLTQ-----TTENLRRVEEDIQOKATGSEVSQRKQOLE-----VEL 1451
QY 620 QKVLLMKNKIDELKQTLILKNVELKHNIHVPSYKQENKQEPYLYLVLLKKEIDKLVFMP 679
DQ 1452 RQVTQMRTEESVR-----YKQSLDDAAKTIQDKNKETERLKR----- 1487
QY 680 KYESLINEEKKNIKTEGQSDNSEPSTEGEITQATTKPCQQAAGSALEGDSVQAQAEQKQ 739
DQ 1488 ----QLIDKETNRK-----CLDENARLQR----- 1509
QY 740 AQPPVPVPPEAKAQVTPPAPVNNKNTENVSKLDYLEKLEFPLNTSYICHKYLIVSHSTM 799
DQ 1510 -----VQYDLQKAN-----SSATETINKLQVQEBELTLRIDY-----ERVS 1546
```

```
QY 800 NEKILKQYKITKEESKISSCDPLDLLFNQIIONIPVMYSMFDSLNNSLSQLFMIEYKEM 859
DQ 1547 QERTVKDQDITR-----FQNSKELQLQKQKVEE 1575
QY 860 VCNLYK---LKONDKIKNLEBAKKVSTSVKTLSSSSMOPLSITPO-DKPEV-----SAND 911
DQ 1576 ELNRLKRTASEDSCKRKKLEEELEGMRRLSK---EQAIKITNLTOQLEQASIVKRSEDD 1632
QY 912 DTSHTSNLNSL-----KLFENILSGKNNIYOELIGOKSSN-FVEKILTKDS 959
DQ 1633 LRQQRDVLGHLREKQRTQOEELRRLSSEVEALRRQLLOEQESVKQAHLENEHFQKAEIDK 1692
QY 960 DTFYNESFTNFVKSADDINSLNDESKRK--KLEEDINKLTKTLQSLFOLYKNYKLIKLER 1017
DQ 1693 SRSLNES-----KIEIERQSLTENLTKEHLMLEEL-----RNLRLLEYDDLRRGRSEADS 1743
QY 1018 LFDKKTGVCKYKMQIKKLTLLKEQLESKLSLNNPKHVL-QNFSVFFNKKKAEIAETEN 1076
DQ 1744 --DKNATILELSQLOISNNRTLEOGLINDLORENERJROETIEKFQKQALEASNRRIQES 1801
QY 1077 TLENTKILLKHYKGLVKYVNGESSPLKTLSESIQOTEDNYASLENFVKLSKLEGLKLDNL 1136
DQ 1802 KNOCTQVQVERESLLVK-----IKVLEQDKARLQRLDELNRAKSTLEAETRVKQRL 1853
QY 1137 NLEKKKL-----SYLSSGLHLLIAELKEVIKKNYNTGNSPSENNTDNNNALESY 1185
DQ 1854 ECEKQIQNDLQWKTQYSRKEEAIKIESEKSEKSNLSRSETERLQAEIKRTEERC 1913
QY 1186 KKFLEPGTDVATVVSSESGDTELOSOPK---KPASTHVCAESN---TITTSQNVDDDEVD 1239
DQ 1914 RRKLEDSTRETQSLETESRYQREIDKLRQRPYSHRETQTECEWTVDTSKLV----- 1967
QY 1240 VIIVPIFGESEEDYDGLQGVVTG-----EAVTPSVIDNLSKIENEYEVL-----YLK- 1287
DQ 1968 -----FDGLRKVTAMQLEYEQCLDKITLQKLLGKKSVSEVASEIQPFLAG 2014
QY 1288 --PLAGVYRSLKKQL-----ENNVTMFWNVK--DILNSRNKKNENKFNWLE 1330
DQ 2015 AGSIAGASAPKESYSLVBAKRRKLISPESTYMLLEAQAATGGIIDPHRNEKLTVDISA 2074
QY 1331 SLDIPYKDLTSSNYVVKDYPKFLNKEKRDKFSSYNIKDSITDINFANDVLGYKILS 1390
DQ 2075 RLIDFDD-----RQIYAA-----EKATGFDPPS 2101
QY 1391 EKYKSDLSIKKYINDQ-----GENEKYLPFLNNI-----ETLY 1425
DQ 2102 KGTVSVSEAIKKNLIDRETGMRLLEAQIASGGVDPVNSVFLP--KDVALARGLIDRLY 2159
QY 1426 KTVNDKIDL---FVHLEAKVLNYYEKSNEVVKELNYLKTQDKLADFKNNNFVGI 1482
DQ 2160 RSLNDPROSKNFVDVFTKKVSYVOLKERCRIEPTHTGLLLSVQ-----KRSMSFQGI 2213
QY 1483 ADLSITDYNHNNLLTKFLSTGMVFENLAKTVLSNLLDGNL 1521
DQ 2214 RQPVTV-----VTELVDSCI-----LRPSTVNELESQGI 2241

RESULT 53
MYSA_DROME
ID MYSA_DROME STANDARD; PRT; 1962 AA.
AC P05661;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin heavy chain, muscle.
GN MHC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
```


AC Q15431; O14963;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Synaptonemal complex protein 1 (SCP-1 protein).
 GN SYCP1 OR SCP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=97224467; PubMed=9119375;
 RA Meuwissen R.L.J., Meerts I., Hoovers J.M.N., Leschot N.J.,
 RA Heyting C.;
 RT "Human synaptonemal complex protein 1 (SCP1): isolation and
 RT characterization of the cDNA and chromosomal localization of the
 RT gene.";
 RL Genomics 39:377-384(1997).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=98037449; PubMed=9371398;
 RA Kondoh N., Nishina Y., Tsuchida J., Koga M., Tanaka H., Uchida K.,
 RA Inazawa J., Taketo M., Nozaki M., Nojima H., Matsumiya K., Namiki M.,
 RA Okuyama A., Nishimune Y.;
 RT "Assignment of synaptonemal complex protein 1 (SCP1) to human
 RT chromosome 1p3 by fluorescence in situ hybridization and its
 RT expression in the testis.";
 RL Cytogenet. Cell Genet. 78:103-104(1997).
 CC !- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.
 CC !- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
 CC !- TISSUE SPECIFICITY: TESTIS.
 CC !- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X95654; CAA64956.1; -
 CC EMBL; D67035; BAA22586.1; -
 CC Genbank; HGNC:11487; SYCP1.
 CC MIM; 602162; -
 CC KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
 CC DNA-binding; Coiled coil.
 FT DOMAIN 12 100 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 107 798 COILED COIL (POTENTIAL).
 FT DOMAIN 117 120 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 679 682 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 880 883 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 961 969 ARG/LYS-RICH (BASIC).
 FT DOMAIN 46 46 L -> F (IN REF. 2).
 FT CONFLICT 106 106 F -> Y (IN REF. 2).
 FT CONFLICT 153 153 F -> C (IN REF. 2).
 FT CONFLICT 161 161 K -> T (IN REF. 2).
 FT CONFLICT 168 168 E -> D (IN REF. 2).
 FT CONFLICT 216 216 N -> S (IN REF. 2).
 FT CONFLICT 225 226 HG -> FE (IN REF. 2).
 FT CONFLICT 230 230 K -> N (IN REF. 2).
 FT CONFLICT 350

FT CONFLICT 360 360 E -> D (IN REF. 2).
 FT CONFLICT 400 401 KN -> NY (IN REF. 2).
 FT CONFLICT 406 406 K -> I (IN REF. 2).
 FT CONFLICT 415 415 K -> T (IN REF. 2).
 FT CONFLICT 449 449 E -> D (IN REF. 2).
 FT CONFLICT 483 510 IQLTAITTSQYYSKEVDLKTLENEK -> YSYCHYHKW
 FT TLPKRGQRKLSSKRE (IN REF. 2).
 FT LTHSKNLSLENK -> YFTLQAQPPPN (IN REF. 2).
 FT CONFLICT 516 528
 FT CONFLICT 549 549 N -> I (IN REF. 2).
 FT CONFLICT 560 560 K -> T (IN REF. 2).
 FT CONFLICT 805 805 E -> D (IN REF. 2).
 FT CONFLICT 941 941 P -> S (IN REF. 2).
 SQ SEQUENCE 976 AA; 114069 MW; 8BA81D042AC2696B CRC64;
 Query Match 3.9%; Score 332; DB 1; Length 976;
 Best Local Similarity 20.7%; Pred. No. 0.00023;
 Matches 236; Conservative 204; Mismatches 354; Indels 344; Gaps 59;
 QY 196 CANDYCOIPF---NLKIRANELD---VLKKLVFGYRKPLDINKDNGVMEDYIKKMK-- 247
 DB 41 CTEDDLLEFPFAKTNLSKNGENIDSPALQKVF---LP---VLEQVNSDCHVQEGLKDS 94
 QY 248 TIEN-----INELTESKTKTDKNKNATKEE-----EKKLYOAO-----YDLST 287
 DB 95 DLENSEGLSRVFSKLYKEAEK-IKKWKVSTEARLOKESKLOENRKIIIEAQRKAIQELQF 153
 QY 288 YNK---QLEEAHNLISVLEKRIDILKKNENIKELDKINEIKNPPPPANGWTPNTLLDK 343
 DB 154 GNEKVSLEEG-----IQENKDLIKENNATRHLCNLLKE-----TCARSAEK 196
 QY 344 NKKIEHEKEIEAKTIKFNIDSLFT-----DPLEYLYREKNKNI-----D 387
 DB 197 TKYEVERETQVYMDLNNIEKMITAHGELRVQAESENRLMHFKLEDEYKIQHLEQ 256
 QY 388 ISAKVETKES-----TEPNYPNGVTYPLSYNDINNALNELNSFGDLNPFDTYKE 438
 DB 257 YKKEINDEKQVSLLIQITKEKENMKDLTFL--EESRDKVNQLE-----EKTKL 305
 QY 439 PSKNIYTDNERKFKNEIKEKIKIEKKTESDKSYED---RSKSLNDITKEYEKLNE 494
 DB 306 QSENLEKQSTEKQHHUTKELEDIKVLSQSVQKALEEDLOIATKTICOLTEKETQME 365
 QY 495 IYDSKFNNDLITNFE-----KMMGKRYSYKVEKLTHHTHTFASVENSKHN--LEKLTKA 546
 DB 366 SNKAAAHFVVVTEFTVCSLEELLRTQORLEKNEQDLKILTMELQKSSSELEMTKL 425
 QY 547 L-KWEDYSLNIVVEKELKYKKNLSKIENIETLVENIKKDEQLEKKTIDENKP 604
 DB 426 TNKVEVEELKAVLGEKETLLYEN-----KQFEKIAEELKGTEQELIGLQAREKEVH 479
 QY 605 DEKILEVSDIVKQVQKVLMLNKIDELKKTOL---ILKNVLEKLNHIVPNSYKQENKQEP 661
 DB 480 D---LEI-QLTAITTSQYYSKEVDL-KTELENEKLNKNTLTSHC---NKLSENKE-- 529
 QY 662 YYLIVLKKEIDKLKVPMPKVESLINEEKNKIKTEGSDNSEPSTGEITGQATTPGQOA 721
 DB 530 -----LTQETSDMTLELNQOEDINNKKQEBRMKLQIENQETETOLNE----- 575
 QY 722 GSALEGDSVQAQQAQKQAPVPVPVPEAKAQVTPPAVNNKNTENV-SKLDYLEKLYE 780
 DB 576 -----LEYVREELKQKR-----DEVCKLQKSENCNLRKQVENKKNYIELOQ 620
 QY 781 FLNTSYICHKYILVSHSTMNKILKQYKITKEESKLSLSCDPLDLFLFNQNNIPVYSMF 840
 DB 621 -----ENKALKKKGTAEKQLNVEI---KVNKLE---LEL----- 650
 QY 841 DSLNLSLSQLFMEI---YEKENVCNLYKLKNDKI---KNLLEAKKVSTSVTLSSSMQ 895
 DB 651 -----ESAKQKFGIETDTYQKEI-----EDKKISEENLLEVEK-----AKVIADAV- 693
 QY 896 PLSLTPQDKPEVSANDDTSHSTNLSNLSKLFENILSLGKNKIYQELIQKSSSENFYKI 955
 DB 1

Db 694 -----KLOKEIDKRCQH-----KIAEMVALMEKKHKHOYDKIEERDSE----- 731
QY 956 LKGDTEVNESFTNFVSKADDSINSLNDESKRKKLEEDINKLKTQLSFDLYNKYKLL 1015
Db 732 -----LGLYSKSEQSSSL-----RASLEIELSNLKAEL-----LSVKKQLEI 769
QY 1016 ERLFDKKTGKYKQMKIKLT-LLKEQLESKLSLNNPKHVLQNFVSFFNKKKEAEIAET 1074
Db 770 ER--EKEE--KLKREAKENTATLKKKKDKK-----TQTF----- 799
QY 1075 ENTLENTKILLKHKYKGLVYKNGESSPLKTLSESIQTEDNYASLENPKVLSKLEGLKD 1134
Db 800 --LLETPEIYWK-----LDSKAVPSQTVSR-----NFTSVDH----- 829
QY 1135 NLNLEKKKLSYL--SSGLHLHIAELKEVIKNKNYTGNSPS-----ENNTDVNNALESYKK-- 1187
Db 830 --GISKDKRDYLWTSAKNTLSTPLP-----KAYTVKTPKPKLQORENLNPIESKKR 882
QY 1188 -----FLPGGTQVATVVSSESGSTLEQ--SQPKKPASTHVGAES-----NTTIT 1229
Db 883 KMAFEFDINDSSSETDILLSMVSE--BETLTKLYRNPNPASHCLVCKTPKKAPSLTT 938

RESULT 56
RASO_SULSO STANDARD; PRT; 864 AA.
AC Q97WH0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 AtPase.
GN RAD50 OR SSO2249.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.,
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006829; AAK42417.1; .
DR InterPro; IPR003439; ABC transportr.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR Pfam; PF02483; SMC_C; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.

FT NP_BIND 30 37 : ATP (BY SIMILARITY).
FT DOMAIN 164 701 COILED COIL (POTENTIAL).
SQ SEQUENCE 864 AA; 101601 MW; 657076AEA9B709FC CRC64;

Query Match 3.9%; Score 331.5; DB 1; Length 864;
Best Local Similarity 20.8%; Pred. No. 0.00021;
Matches 230; Conservative 200; Mismatches 321; Indels 357; Gaps 55;

QY 192 LNDVCANDYCOIPENLKIR-----ANELDVLKLVFGY-----RKPLDNI-----KDNVG 236
Db 8 LTNFSLSEHSIQPMGEINLVIGONGACKSSIIDGIVFSLFRHSRGNNDNLIRKGSNRG 67
QY 237 KMEDYIKRKNKTIENINELIEESKKTIDKN-----KNAT--KEEEKKKLYQAYDLSYKN 290
Db 68 SVTLYSNSKDKIEIRDSTTDEDLIRNQFPPIARSATVVSNEIEKILGIDKIDIAL-- 124
QY 291 QLEAHNLISVLEKRIDTLKKNENIKELLDKINEIKNPPSPANSNGTPTNLLDKNKKIEH 350
Db 125 -----STIIVRQGEID--KILENFQETMGKI-----LKLEI 154
QY 351 EKEIKEIATAKTIKFNIDSLFTDPLEYVLRKKNKIDISAKVETKESTEPNPGVTYP 410
Db 155 EKLDSRGPIVEF-----RKNLE----- 172
QY 411 LSYNDINNALNELNSFGDLINPFDTYKPSKNITYTDNRKKFNEIKKIEKKIESD 470
Db 173 -----NKLRELDRIEQDYNFKKTV-----EKARVLELKK-----DREKLEDE 212
QY 471 KKSIEDRSKSLNDITKEYEKLLNELYDSKFNNDILTNFEKMMGKRYKYKVEKLTHHN-T 529
Db 213 IKNEIKRIKIDKQDFDEYK-----KRNOYLKLTTLTKI-----KEGELNELRS 257
QY 530 FASYENSKHNLEKLTALKYMEDYSLRNIVVEKELKYYK-NLISKIENEIETLVENIKKD 588
Db 258 IEELRKOTENWDQLEKEINELE--NLNRI-----KLKPEKYEVLAKSHTSEMANVINLEKE 311
QY 589 EQLFEKKI--TKDENKPKDEKILEVSDIVKQVQVLLMKNKIDELK--KTQLILK----- 639
Db 312 IEE-YEKAIRRKELEP--KYLKYE-----LERKLEELQPKYQVLLKLSLDLS 358
QY 640 --NV-----ELKHNHVPNSYKQENKOEPPYLIYVLKKEIDKLKVPMPKVESLINEE 688
Db 359 KNLKERLEKDASELSNDIDKVNLSLEQVEE-----TRKKOLN-LRAQLAKVESLISEK 411
QY 689 KKNIKTEGQSDNSPSTEGEITGOATKPGQAGSALSGDSVQAGAQKQKQAPVPVPV 748
Db 412 NEII-----NNISQVEGE-----TCP--VCGRPLD-----BEHKQK-----II 442
QY 749 PEAKAQVPTPPAPVNNKNTENVSK-----LDYLEKLYEFINTSYICHKYILVSHSTMNEKIL 804
Db 443 KEAKSYILQLELKNLEELKKTITNELNKTREYRRLS-----NKKASYDNVM 491
QY 805 KOYKITEEESKLSLSCDPLDLFIQNNIPVMSWFDLSNLSQLFMEIYEKEMVCNLY 864
Db 492 RQLAKLNEETIENLHS--EIESLKNIDEEI-----KKINBEVKELKLYEEFMRLSKY 541
QY 865 KLKDNDAIKNLLEAKKYSTSVKTLSSSSMOPLSLTPDQKPEVSANDDTSHSTNL--N 920
Db 542 TKEELDKRVKLDKMKKKEIE-----KEMRGL-----ESELGLDKRALESKILDLEN 591
QY 921 NSLKFNILSLGKNKNIYQELIG-----QKSSSENFYEK--ILKSDPTFYNESEFTNFYKS 973
Db 592 KRVLDE-----MKKKKGILEDYIRQVKLQEEVKNLREEVNIQFDENRYNE-----LKT 642
QY 974 KADDIN-SLNDESKRK-KLEEDINKLKTLOLSFDLYNKYKLERLDFDKKTV----- 1025
Db 643 SLDAYNLSLKEKENRKSRIEGELESLEKIDIEISNRIANYELQLK--DREKIINAKNL 699
QY 1026 -----GKYKMOIKKLTLLKEQLESKLSLNNPKHVLQNFVSFFNKKKEAEIAETNTL 1078
Db 700 EKIRSGALGERKLOSYIIMTTKQLIENNNDI-----ISKFDUSI 738
QY 1079 ENT--KILLKHYG-----LVKYNGESSPLKTL-----EESIQTED 1114

Db 1067 ERVKRELEENVRKKAETSSLSKLEDEQNLSQLOKTKIKELQARIEELEBELEARNAR 1126
QY ~, 593 --REK---KITDENKPEKILEVDIVKVQVQKVLNKNKIDELKKTQILKNVELKHNI 647
Db 1127 AKVEQRAELNDELGERLDEAGATSAQIE--LNKKREALLKIRDLLEASIQHEA 1184
QY 648 HVPNSYK--QENKQBPYILVLKELDKLVKMPKVESLINEEK-----690
Db 1185 QISALRKKHQDAANE-----MADQVDQLQ-----RVKSKLEKDKDLKREMDLLESQWTH 1234
QY 691 NIKTEGQSDNSPSPTEGITG--QATTKPGQAGSALGDSVQAAQ-----EQKQAOPP 743
Db 1235 NKNKGCSEKVMKQFESQMSDLNARLEDSORSINELQSKSRQLENSDLTRQLEDAEHR 1294
QY 744 VVPVPE---AKAQVPTPPAPVNNKNTENVKLDYLEKLYEFLNTSYICHKYILVSHSTMN 800
Db 1295 VSVLSKERSQLSQLESDARRSLEETRSKL-----QNEVRNMHADN- 1337
QY 801 EKILQYKTKTEESKLSKSCDPLDILFNIONNIPVWYSWFDS-----LNNLSQLPM 852
Db 1338 DAIREQ--LEEQESK---SDVQORLSKANNEIOQWRSKFSEGANRTTEEDQKRKLIG 1392
QY 853 EYKEMVCNLYKLDN--DKIKNLE-----EAKKVSTSVKTL-----SSSS 893
Db 1393 KLSEAEQTEAANAKCSALEKAKSLQOBLDMSIEVDNANASVNMCKKQRAFDKTAE 1452
QY 894 MQLSLTPQDKPEVANDTSHSTNL---NNSLKLFIENLSLG-----KNKNLYQBELI --- 943
Db 1453 WQAKVNSLQSELENSQESRGVSAELYRIKASIEYQD--SIGALRRKNKNLADEIHDIT 1510
QY 944 -----GOKSSE-----NYYE 953
Db 1511 DOLSEGRSTHELDKARRLEMEKEELOALAEAGALEQBEAKVMRAQLEIATVRNEID 1570
QY 954 KILKSDTFYNESFTNFVKSKADDINSLEDSK-----RKKLEEDINKLTKLQLSF 1005
Db 1571 KRIQKEEEDFNTRRNHQRALSMQASLEAEAKGAKADAMRIKKLEQDINE-----LEVAL 1626
QY 1006 DLYNRYKLEKLEKLFKTKVTGKYMQIKKL-----TLIKE 1040
Db 1627 DASNRGAEME-----KTVRYQOQIREMQTSIEEQORDEARESYNMAERRCTLMSG 1680
QY 1041 QLESKLNLSNPKHVLFQNFVFFNKKEAEIAETENTLENKILKHKYGLVKVYNGBSS 1100
Db 1681 EYEELRAALEQAERA-----RRASDNLADANDRYNELTSQVSSVQGGKRLKLEGDIN 1732
QY 1101 PLKT-----LSPESITQEDNYASLENFVLSKLECKLD- 1134
Db 1733 AMQTDLMHGLKGADECKKAMADAARLADLRAEQDHSNQVE--KVRKNLESQVKEF 1790
QY 1135 -----NLNLKPKKLSVLSGLHHLIAEL-----KEVIKNKN-----YT 1167
Db 1791 QIRLDEAEASSIKGGCKMIQKLESVHELEALDNEQRRHAETQKNMRKADRLKELAFQ 1850
QY 1168 GNSPSENN-----TDVNNALESYKKFLPEGTQVATV-----VSSSG 1203
Db 1851 ADEDRKNOERLQELIDKLNAKIKTFKQVVEEAETIAINLAKRYKAQHELEAEERADTA 1910
QY 1204 SDTLFOSQPKKPASTHVCAESNFIITS 1230
Db 1911 DSTLQFRKFRSSSVQSRSSVSAS 1937

RESULT 58
MYSS_CHICK
ID MYSS_CHICK STANDARD; PRT; 1938 AA.
AC P13538; Q13228;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, adult.
OS Gallus gallus (Chicken).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=white leghorn; TISSUE=Pectoralis muscle;
RA Chao T.H., Bandman E., Moore L.;
RT "Cloning, nucleotide sequence and characterization of a full-length
RT cDNA encoding the myosin heavy chain from adult chicken pectoralis
RT major muscle.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-205.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041767; PubMed=1939027;
RA Hayashida M., Maita T., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: I.
RT Sequence of the amino-terminal 23 kDa fragment.";
RL J. Biochem. 110:54-59(1991).
RN [3]
RP SEQUENCE OF 206-636.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041768; PubMed=1939028;
RA Komine Y., Maita T., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: II.
RT Sequence of the 50 kDa fragment of subfragment-1.";
RL J. Biochem. 110:60-67(1991).
RN [4]
RP SEQUENCE OF 637-837.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041769; PubMed=1939029;
RA Maita T., Miyaniishi T., Matsuzono K., Tanioka Y., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: III.
RT Sequence of the 22 kDa fragment and the alignment of the 23 kDa, 50
RT kDa, and 22 kDa fragments.";
RL J. Biochem. 110:88-74(1991).
RN [5]
RP SEQUENCE OF 838-1938.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041770; PubMed=1939030;
RA Maita T., Yajima E., Negata S., Miyaniishi T., Nakayama S., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: IV.
RT Sequence of the rod, and the complete 1,938-residue sequence of the
RT heavy chain.";
RL J. Biochem. 110:75-87(1991).
RN [6]
RP PRELIMINARY SEQUENCE OF 1-808
RX MEDLINE=87092420; PubMed=3467365;
RA Maita T., Hayashida M., Tanioka Y., Komine Y., Matsuda G.;
RT "The primary structure of the myosin head.";
RL proc. Natl. Acad. Sci. U.S.A. 84:416-420(1987).
RN [7]
RP SEQUENCE OF 842-1270.
RX MEDLINE=90121764; PubMed=2610940;
RA Watanabe B.;
RT "Complete amino-acid sequence of subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:1027-1034(1989).
RN [8]
RP SEQUENCE OF 852-1108.
RX MEDLINE=89374803; PubMed=2775482;
RA Watanabe B.;
RT "Amino-acid sequence of the short subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:549-558(1989).
RN [9]
RP SEQUENCE OF 1145-1270.
RX MEDLINE=89228549; PubMed=2713098;
RA Watanabe B.;
RT "Amino-acid sequence of the hinge region in chicken myosin
RT subfragment-2.";
RL Biol. Chem. Hoppe-Seyler 370:55-61(1989).
RN [10]

RP SEQUENCE OF 1957-1938 FROM N.A.
RX MEDLINE=87217964; PubMed=3034534;
RA Motilary D.M., Barringer K.J., Dodgson J.B., Richter H.E.,
RA Young R.B.;
RT "Genomic clones encoding chicken myosin heavy-chain genes.";
RL DNA 6:91-99(1987).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 1-843.
RX MEDLINE=93303624; PubMed=8316857;
RA Rayment I., Rypniewski W.R., Schmidt-Base K., Smith R.,
RA Tomchick D.R., Benning M.M., Winkelmann D.A., Wesenberg G.,
RA Holden H.M.;
RT "Three-dimensional structure of myosin subfragment-1: a molecular
motor.";
RL Science 261:50-58(1993).
CC -!- FUNCTION: MUSCLE CONTRACTION. MYOSIN IS A PROTEIN THAT BINDS TO
CC F-ACTIN AND HAS ATPASE ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U87231; AAB47555.1; -;
DR EMBL; M16557; AAA48970.1; -;
DR PIR; PX0050; PX0051.
DR PIR; A26821; A26821.
DR PIR; S02082; S02082.
DR PIR; S04501; S04501.
DR PIR; S05515; S05515.
DR PDB; 2MIS; 1I-JAN-97.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Phosphorylation; Acetylation;
KW Calmodulin-binding; Multigene family; 3D-structure.
FT INIT_MET 0
FT DOMAIN 0
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IQ.
FT DOMAIN 838 840 HINGE.
FT DOMAIN 841 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT MOD_RES 1 1 ACETYLATION.

FT	MOD_RES	35	35	METHYLATION (MONO-).
FT	MOD_RES	130	130	METHYLATION (TRI-).
FT	MOD_RES	551	551	METHYLATION (TRI-).
FT	MOD_RES	755	755	METHYLATION (MONO-).
FT	MOD_RES	697	697	ALKYLATION (SH-1).
FT	MOD_RES	707	707	ALKYLATION (SH-2).
FT	CONFLICT	907	907	C -> Q (IN REF. 7 AND 8).
FT	CONFLICT	980	980	L -> F (IN REF. 1).
FT	CONFLICT	1343	1343	E -> D (IN REF. 5).
FT	CONFLICT	1545	1545	S -> A (IN REF. 5).
FT	CONFLICT	1796	1797	HV -> QL (IN REF. 5).
FT	CONFLICT	1830	1830	S -> A (IN REF. 5).
FT	CONFLICT	1863	1863	I -> V (IN REF. 10).
FT	CONFLICT	1929	1931	IHG -> FH (IN REF. 10).
SQ	SEQUENCE	1938 AA; 223013 MW; EDD01CEA2681E10F CRC64;		

Query Match 3.9%; Score 331; DB 1; Length 1938;
Best Local Similarity 19.1%; Pred. No. 0.00053;
Matches 241; Conservative 233; Mismatches 434; Indels 354; Gaps 56;

QY	237	KMEDYIK--KNKKTININELIEESKTTIDKNKNATKEEEKK--LYQAQYDLSIYNKQL	292
DB	837	KIKPLKSAESEKEMANKKEFEKTKELAKSEAKKELEKMMVYLQEKNDLQ---	QV 893
QY	293	EEAHNLISVLEKRIDTLKKN-----NIKELDKINEIKNPPSPANGTNPNTLLDKNKI	347
DB	894	QAEADSLADAERCQDLIKTKLEAKIKVEERADEEE--INA-----ELTAKKRKL	945
QY	348	EEHEKEIKEIAKTIFNIDSLTDPLELYLRKNKNIDISAKVE-TKESTEPNEYPNG	406
DB	946	ED---ECSELKKDI-----DLELPL-----AKVEKEKHATE-----	974
QY	407	VTPYLSYNDINNALNELNSFGDLPDYTKPSKNIYTDNERKKFINIKIEKKIEKK	466
DB	975	-----NKVKNLTEMAVLDTIAKL--TKE-----KKALQEAHQ-TLDDLQ	1013
QY	467	ISDCKSYEDRSKSLNDITKEYELNIEYDSKFNNDLTNFEKMGKRYKYVE---K	523
DB	1014	VEEDK-----VNTLTAKTKLEQQVDD--LEGSLEQEKKLMDLERAKRKLEGD	1062
QY	524	LPHHTFASYENSKNLEKLTALKY-----MEDYSLRNIVVEKLYKNLISKIE	575
DB	1063	LA-HDSIMDLENDKQOLDEKLAKKDFEISQISQKTEDEQALGMLOKKIKELQARIE	1121
QY	576	NEIETLVENIKKDEQLEFKEKITDENKDPKILEYSDIVKVQVQVLLMNK--IDEL	633
DB	1122	EEIE--AERTSRAKAEKHRADLSRELEEFISERLEEAGGATAAQIE---MNKK	1175
QY	634	TQILKNVELKH-----NIHVPNSYKQENKQEPYIYLVLKKE	670
DB	1176	MRRDLEEATLQHEATAAALRKKHADSTABELGEQIDNLQRVKQKLEKESE---	1229
QY	671	ID-----KLKVPKPVESLINEEKKNIKTQSDNSEPSTEGITQATKPGQ	720
DB	1230	IDDLASNMESVSKAKANLEKMCRTLEDQLSEIKTEEQNRMINDLNTORARLTGT	1289
QY	721	AGSALGDSVQAQAEQKQAPVPVPVPEAKAQVTPPAPVNNKNTENYSKLD-YLE	779
DB	1290	SQAEEKDALISQLSRGQ-----GFTQOIEELKRLHLEE--	1323
QY	780	EFLNTSYCHKYLVSHSTPMNEKILKYKITKEESKLSLSCDPLDLLFNQNNIPV	839
DB	1324	EIKAKNALAHALQSAHDC--ELLREQYE--EEQEA-----	1356
QY	840	FDLSLNNLSOLFWEI-----YEKEMVCNLYKLD-NDKIKNLEBAKVKSVTLSS	892
DB	1357	-GELQALSKANSEVAQWTKYETDAIQRTTELEAKKLAQRLQDAEEHVAVNA-K	1414
QY	893	SMQPSLTPODKPE--VSANDTSHSTNLNNSLKFENILSLGNKNYIYQEL----	944
DB	1415	SLEKTKQRLQNEVEDLWVDVRSNACALDKKQNFDKILAENKQK--YEETQTE	1472
QY	945	QKSS-----ENFYEKILKSDTTFYNESFTNFVKSADDINSLND-----	983

Db 1473 OKESRSLSTELFKMKNAYEGLDHLTKREN-----KNLQOEADLTETQIAEGGKAVHE 1527
Qy 984 -ESRKKLEEDINKL-----KTLQSLDLYNKYKXLER-LFDKKTG 1026
Db 1528 LEKVKHVEQKSELQASLEAEASLEHEGKIIRLQLEL-NQIKSEIDRKIAEKDEID 1586
Qy 1027 KYKM-QIKKLTLLKEQLESKLNPNKPHVLQNFVSFFNKKKEAEIAETENTLENTKILL 1085
Db 1587 QLRNHLRIVESMOSTLDAEIRS-----RNEALRLKKMEGDLNEMEIQLSHANRMA 1638
Qy 1086 KHYGLVKYNGESSPLKTLSEESIQEDNVASLENFKVLSKLEGLKDNINLEKKLSY 1145
Db 1639 AEAKNLRLNTQG-----TKDQIHLDDALRTQED-----LKEQVAVYERRANL 1682
Qy 1146 LSSGLHLIAELKEVINKNYTGNPSNNTDV-----NNALSYKKFLPEGTDVATVV 1199
Db 1683 LOAEVEELRGALQETRSKVAEQLDATERVOLLHTQNTSLNTKKLE--TDIVOIQ 1740
Qy 1200 SEGSSTLEQS-----QPKKPASTHWGAESNTITTSQ-----NVDDVDV---- 1240
Db 1741 SEM-EDIOEARNAEAKKAITDAAMMAEELKKEQDTSAHLEKMKNMDOQTVADLHVRL 1799
Qy 1241 -----IIVPIGESEEDYDDLGOVVTGEAVTSPSIDNLSKIENE 1280
Db 1800 DEABQALKGGKQLOKLEARVRELEGEVDEQKRSAAEAVKGVKRYERRVKELTYQCEED 1859
Qy 1281 YE-VLYLKPLAG-----VYRSLLKQLENNVMTFNVVK-----DILNSRF 1319
Db 1860 RKNILRLQDLVKLOMKVYSKQAEAEELSNVLSKFRKIQHELEAEERADIAESQV 1919
Qy 1320 NK 1321
Db 1920 NK 1921

RESULT 59

TANA_XENLA STANDARD; PRT; 1744 AA.
ID TANA_XENLA
AC Q01550;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tabin.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUP-Tadpole head;
RX MEDLINE=92398961; PubMed=1524825;
RA Hemati-Brivanlou A., Mann R.W., Harland R.M.;
RT "A protein expressed in the growth cones of embryonic vertebrate
neurons defines a new class of intermediate filament protein.";
RL Neuron 9:417-428(1992).
CC -1- TISSUE SPECIFICITY: GROWTH CONES OF EMBRYONIC VERTEBRATE NEURONS.
CC -1- DEVELOPMENTAL STAGE: IS EXPRESSED IN THE NEURULA AND PERSIST
CC DURING EMBRYOGENESIS IN THE BRAIN, CRANIAL NERVES, AND SPINAL
CC CORD.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; M99387; AAA49966.1;
DR PIR; JH0720; JH0720.

DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone.
FT DOMAIN 1 12 HEAD.
FT DOMAIN 13 314 ROD.
FT DOMAIN 315 1744 TAIL.
FT DOMAIN 8 48 COIL 1A.
FT DOMAIN 49 60 COIL 1B.
FT DOMAIN 61 156 COIL 1C.
FT DOMAIN 157 179 LINKER 12.
FT DOMAIN 180 193 COIL 2A.
FT DOMAIN 194 199 LINKER 2.
FT DOMAIN 200 314 COIL 2B.
SQ SEQUENCE 1744 AA; 199561 MW; 6502EAC9FE6C4E93 CRC64;
Query Match 3.9%; Score 330.5; DB 1; Length 1744;
Best Local Similarity 18.0%; Pred. No. 0.00049;
Matches 305; Conservative 311; Mismatches 574; Indels 509; Gaps 79;
Qy 132 VRNYLLTIKELKYPQLDNLNMLTLCNIIHGFKYLDGYEEINELLYKLNFPDLLRAK 191
Db 1 MEGYLASV-----SLGEESTQMWSLNKRLRAYLSRVKALEEENELLRK---EIHSLRSS 51
Qy 192 LNDVCANDYCOIPFNLRIRANELDLVKLVFGYRKPL--DNIKDNVGMEDYIKNNKTI 249
Db 52 KSERC-----WKKKHHEEMMKLRDALDDGHEMVAQEMVRSIYEIEFVK--QRCL 101
Qy 250 ENINELIEESKTTDKNNATKEEKKKLYOAYDLSYNNKQLEEAHNLISVLEKRI--DT 308
Db 102 EE-KOAREDAKELSESKELEETRAQIW-----LKER---LGQLEAELEDI 145
Qy 309 LKKNENIKELLDKINEIKNPPANSNTPTLLDNKKNIEEKEIKEIAKTIK--FNIDS 367
Db 146 LRDEEEKALM-----EELIASFSORLENFRVAP 174
Qy 368 LFTDPLEYLYREKNNKIDISAKVETKESTEPNEYPNGVTYPLSYNDINNALNELNSFG 427
Db 175 VAFKPVVDYAR-----KLSEIWQGAVEEYKSEVS-----LEAGLSE----- 213
Qy 428 DLINPFDYTKEPSKNITYDNERRKFFINEIKEK-----IKIEKKKIES-DKKSYEORSKSLN 482
Db 214 -----SKENLRKVLLENKONRLILQSLDKELVSLKMRKEALEDLLSKOMQOEKEEEE 265
Qy 483 DITREKLLNEIYDSKFNNNIDLTNPEKMMGKRYSYVEKLTHTHTNTPASYENSKHNLEK 542
Db 266 KLOKKALEAEKQDLRCQIAEVLEDRQOLMLKMSLEVATYRSLEA-----ES 317
Qy 543 LTKALKYMEDYSLRNIVVEKELKYYKNLSKIENIEFTLVENIKKDEQLEKFKITKDN 602
Db 318 TRIYTDYRGSYTFNDMSLE-----HNNVRRR---QSEDTKRTVSKDHRQSYKKQIGDN 369
Qy 603 ---KPD-----EKILEVSDIVKVOVKVLLMKNKIDELKKTQLILK 639
Db 370 ELQPSLNNFTSVKSSAVPVRTSPVTKFQKVSQGLKYTKAPQVKVQTVSTVYS 429
Qy 640 NVELK-----HNIWPNYSKOENK-QEPPYL--IVLK 668
Db 430 NLEHTTSGDAFRAQVETRTKDEQVKKDALGLNDLNKNTGFKKEKDIQKPGFMHVVS 489
Qy 669 KEIDKLVFMPKVESL-----INEEKNIKT---EGQSDNSEPSTEGEITGOATKPGQ 719
Db 490 KSVSSTHEKVPIDPLESALSKEEDLSVSVSTFNAGQSSNLEAIK--DVLGECLENL 547
Qy 720 QAGSALEGDSVQAQAEQKQAPVPVPVPEAKAQVTPPAPVNNKNTENSKDYLEKLY 779
Db 548 QNEIAFEKESPGTNA-----AADPIEEVISE-----VSQTVHFKEQE-LSNLL 591
Qy 780 EFLNTSYICHRYILVSHSTMNEKILKQYKTKKEESKLSSC--DPLDLLLFIQNNIPVY 837
Db 592 EIENT-----HE---NH-----VQDQTAQFNSCQDGHDRASTLENNEPDVQ 630
Qy 838 SMFDSLNNLSOLFMEIYEKEMVNCNLYKLNKNDKIKNLLEEAKKYSTVSKVTLSSSSMQPL 897

```
Db 631 QVIRLESN-----EIKESKIP-----SDN-----TEEAIIISKRKVLENEYIPV 672
QY 898 SLTPODKPEVSANDOTSHSTNLNLSKLPEN-----ILSLGKN--KNIVQELIGQKSS 948
Db 673 S--KODLTFEFTSHLENDSESSQSFOSKUFENKSTEDQLITNLKSNTOENIFOS--NQEHL 728
QY 949 ENF-YEKILKDSDTFYNESFTNFVKS-----ADDINSLNDES-----985
Db 729 ENLEFDSVVPDIVKFWYQENLLBEEENVYGDGELVQMATDENIINQSSDQLLLSDHSHH 788
QY 986 -----KRRKLEEDINKKTLQLSFDLYNKYKLK--LERLFDKKTVGYKMQOI 1032
Db 789 EETKTSESTAVENHMESEHAEDVKSSEIPVEISENSVSEIIEHISDVEEDT-----841
QY 1033 KKLTLKQLESKLANPKHVLQNFVSFFNK-----KKEAEIAE---1073
Db 842 -KQAFEDERVGOINNONNOESTVDLDGVSYSQEEENSQLEEDSVSISEQIERDKFEINEQEC 900
QY 1074 -----TENTLENTKILKHYKGLVKKYNGESSPLKTLSEESI 1110
Db 901 LKSDOIREFDTEEDVHQVDFPMQSFQSFEREVGQLNNIKQEVGY-----LQNYDEDSF 953
QY 1111 QTEDNYASLENPKVLSKLEGLKDNLNLEKKKLSYSSGLHHLIAELKEVINKKNYTGNS 1170
Db 954 QNNDPEQLESCDL--QEQKIK---LEEEN-----QLSENECQNFQGGND 993
QY 1171 PSENNTDVNNALSYKKLPCTGVATVVS-----SGSDTLEQ 1209
Db 994 IEE-----FSQCYDTEICQETIGNQVSAQLCESDINODKLSMEDEEQ 1039
QY 1210 SQPKKPASTHVGAESNTITTSN-----VDDEVDVVIIPFGESEEDY-----DILGQVVT 1261
Db 1040 NNPETEDNIGLEQSDQENTRSNEGTKEFSQECDVVFPEDMSDKSEYSGQOEDLDKQVT 1099
QY 1262 GBAVTPSVIDNTLSKIENEYEVLYLKPLAGVYRSKQLENNVMTFNVNVDILNSRPNK 1321
Db 1100 DFLNEQANNOLL--EKEEVLH--HADDORS-----VNDEITIDEKLSE 1140
QY 1322 RENFNKVLSDLPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDIDINFAND 1381
Db 1141 R-----IIDNELATV-DVNES-----LAANKQVDLFTDEY-----AVDDNYGMQDD 1181
QY 1382 VLGYKYLSEKYKSDLDSTIKKY-INDKOGENEKYLPLFNNI-----ETLYKTV--NDKI 1432
Db 1182 DSGQYQTKREDLFVDCGNIIETIEIOOTSLNQEICERVQNVNVEDISGEAKNESVEMNDVV 1241
QY 1433 DLFVTHLEAKVLYNYEKSNEVVKIKELNYLKTIOD-----KLADFKN--NNFVGIAD- 1484
Db 1242 DLVP---BAKV---RGDEQISPLQDEKLN-LETMEDTKDNDQCLKEKENETEYIEVTDS 1294
QY 1485 --LSTDYNNH---NLLTKFLSGVVF-ENLAKTVLSN-----LLDGNLOGMLNISQHC 1532
Db 1295 POFATYLDHSHDAGRELTVPQNSANLOFCENPTKTLIAHIEYETVADSLES-----TEEQ- 1349
QY 1533 VKKQCPQNSGCPRHLDERECKCLNYKQEGDKCVENPNPTCNENGGCDADAKTEEDS 1592
Db 1350 -----VQETER-----IPFKPEDSK-MENENSESESVDSQEISLSNHSKSEE 1390
QY 1593 GSNKKKITCECTKPDSPYL 1611
Db 1391 FEISKDYQLEOTLPDWTPL 1409
```

Search completed: March 31, 2003, 07:21:27
Job time : 124 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 07:20:46 ; Search time 38 Seconds
(without alignments)
4146.429 Million cell updates/sec

Title: US-09-269-874A-3

Perfect score: 8424

Sequence: 1 MKLIFLCSFLFFIINTQCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8424	100.0	1639	2 S05603	major merozoite su
2	8191.5	97.2	1640	2 A24594	probable major sur
3	7965	94.6	1631	1 SAZOK1	major merozoite su
4	5136.5	61.0	1726	1 SAZQGM	major merozoite su
5	5121.5	60.8	1726	2 A54948	major merozoite su
6	5065	60.1	1701	2 A54498	major merozoite su
7	5053	60.0	1701	2 A26868	major merozoite su
8	3093.5	36.7	1751	2 A45604	major blood-stage
9	2999.5	35.6	1086	2 S16752	major merozoite su
10	2359	35.1	1726	2 A39401	merozoite surface
11	2829.5	33.6	1060	2 S06286	major merozoite su
12	2548.5	30.3	1772	2 A45532	major merozoite su
13	2509	29.8	1785	2 A45546	major merozoite su
14	2146	25.5	651	2 S47282	merozoite surface
15	2101	24.9	400	2 A45545	major merozoite su
16	1187	14.1	233	2 A25814	merozoite surface
17	1135	13.5	680	2 A28121	glycoprotein 185 -
18	1096.5	13.0	281	2 B54498	major merozoite su
19	1021.5	12.1	636	2 A45949	merozoite surface
20	815	9.7	168	2 A25522	major merozoite su
21	710.5	8.4	146	2 S03290	glycoprotein, 190K
22	659.5	7.8	138	2 P00125	major merozoite su
23	605.5	7.2	1939	2 T18372	repeat organellar
24	605	7.2	2401	2 T28676	rhostry protein -
25	568.5	6.7	2269	2 T28677	rhostry protein -
26	559.5	6.6	1979	2 C71622	hypothetical prote
27	529	6.3	115	2 B25120	major merozoite su
28	521.5	6.2	3394	2 T18501	hypothetical prote
29	507	6.0	2166	2 G70163	hypothetical prote

hypothetical coile
merozoite surface
transport protein
ORF MSV156 hypothe
reticulocyte-bindin
hypothetical prote
myosin-like protei
merozoite surface
merozoite surface
reticulocyte-bindin
serine/threonine-s
hypothetical prote
centromere protein
hypothetical prote
hypothetical prote
interaptn - slime
major merozoite su
surface-located me
conserved hypothet
lmp1 protein - Myc
hypothetical prote
myosin heavy chain
hypothetical prote
DNA polymerase - m
hypothetical prote
myosin heavy chain
hypothetical prote
major merozoite su
phosphatase (acid
giantin - human
RAD2 endonuclease
hypothetical prote
major merozoite su
giantin - human
major merozoite su
hypothetical prote
hypothetical prote
major merozoite su
hypothetical prote
ovt1 protein - nem
myosin heavy chain
major merozoite su
hypothetical prote
major merozoite su
myosin-like coiled
probable membrane
probable nuclear p
RAD50 protein - ye
kinesin-related pr
probable membrane
hypothetical prote
nuclear migration
protein g377 - mal
hypothetical prote
chromosome segrega
hypothetical prote
major merozoite su
hypothetical prote
hypothetical prote
hypothetical prote
surface-located me
chromosome assembl
major merozoite su
DNA-directed RNA p
hypothetical prote
mature-parasite-in
hypothetical prote

103 365.5 4.3 2380 2 E71604
104 365 4.3 1690 2 T13030
105 364 4.3 1447 2 F82909
106 364 4.3 1642 2 T08880
107 362.5 4.3 1712 2 C71618
108 361.5 4.3 2244 2 F90563
109 361 4.3 3660 1 S02041
110 360.5 4.3 1714 2 F16009
111 360.5 4.3 2346 2 T13829
112 360.5 4.3 2708 2 T09079
113 359.5 4.3 4981 2 T18489
114 359 4.3 1271 2 A45555
115 358.5 4.3 2094 2 S31124
116 357.5 4.2 1817 2 H71611
117 356.5 4.2 971 2 A70179
118 356.5 4.2 1294 2 T18473
119 356 4.2 2599 2 F90608
120 355 4.2 1173 2 T43527
121 354 4.2 1997 2 F17607
122 353.5 4.2 2285 2 T12796
123 353 4.2 1169 2 A64505
124 352.5 4.2 1163 2 G97236
125 351.5 4.2 1935 1 S06006
126 351 4.2 2253 2 T30336
127 351 4.2 5005 2 F82884
128 350.5 4.2 1177 2 B75150
129 350 4.2 1802 2 G71616
130 349.5 4.1 1935 1 A37102
131 349.5 4.1 1940 2 A29320
132 349 4.1 1939 2 I48175
133 348 4.1 1532 2 T18438
134 348 4.1 2657 2 T18497
135 347 4.1 1225 2 A49464
136 346 4.1 1306 2 T28313
137 346 4.1 1819 2 A71928
138 346 4.1 1927 2 A59236
139 346 4.1 2819 2 T05080
140 345.5 4.1 1183 2 F90559
141 345.5 4.1 1927 2 G64585
142 345.5 4.1 2649 2 A40937
143 345 4.1 1313 2 A48467
144 345 4.1 1620 2 S61535
145 345 4.1 1817 2 D46762
146 345 4.1 1939 1 A76762
147 344.5 4.1 978 2 F07387
148 344.5 4.1 1830 2 E82909
149 344.5 4.1 1934 2 I48153
150 344 4.1 2136 2 A05037
151 343.5 4.1 1338 2 T18416
152 343.5 4.1 1902 2 C97702
153 343 4.1 1236 2 T18459
154 343 4.1 1367 2 T18466
155 342 4.1 2052 2 C97038
156 341.5 4.1 1269 2 F84730
157 341.5 4.1 1938 2 I49464
158 341.5 4.1 1979 1 S03166
159 341.5 4.1 2437 2 T18482
160 339.5 4.0 1787 2 G97222
161 339 4.0 100 2 C44866
162 338.5 4.0 1957 2 A45627
163 338.5 4.0 2364 2 I40884
164 338 4.0 1169 2 H70178
165 338 4.0 1938 1 S06005
166 337.5 4.0 1051 2 T18351
167 337.5 4.0 1164 2 T24806
168 336.5 4.0 1427 2 S22695
169 335.5 4.0 1272 2 C90593
170 335 4.0 1955 2 T30934
171 334.5 4.0 2677 2 A38194
172 334 4.0 1392 2 A43336
173 333.5 4.0 2178 2 S55805
174 333 4.0 1538 2 T29095
175 332.5 3.9 1935 2 A59286

hypothetical prote
microtubule bindin
hypothetical prote
NMDA receptor-bind
hypothetical prote
hypothetical prote
dystrophin, muscle
Ser/Thr protein ki
tpr homolog - frui
probable chloroqui
hypothetical prote
glutamate rich pro
tpr protein - huma
probable secreted
exodeoxyribonuclea
hypothetical prote
ABC transporter pe
sp8 protein - fliss
DNA helicase II BR
probable transglyc
P115 homolog - Met
ATPase involved in
myosin beta heavy
nuclear/mitotic ap
hypothetical prote
chromosome segrega
hypothetical prote
myosin beta heavy
myosin heavy chain
myosin heavy chain
hypothetical prote
hypothetical prote
ORF MSV152 probabl
cag island protein
embryonic muscle m
probable chloroqui
conserved hypothet
cag pathogenicity
bullous pemphigoid
myosin heavy chain
nucleotide-binding
hypothetical prote
myosin alpha heavy
conserved hypothet
myosin heavy chain
hypothetical prote
cell surface antig
hypothetical prote
hypothetical prote
phage-related prot
probable myosin he
alpha cardiac myos
myosin heavy chain
hypothetical prote
hypothetical prote
major merozoite su
myosin heavy chain
cytotoxin L - Clos
exodeoxyribonuclea
myosin alpha heavy
lmp1 protein - Myc
hypothetical prote
resin - human
hypothetical prote
myosin-like protei
desmoplakin I - hu
microtubule-vesicl
alpha-toxin - Clos
cardiac muscle fac
myosin heavy chain

hypothetical prote
purine NTPase [imp
myosin heavy chain
dystrophin, muscle
tanabin - African

ALIGNMENTS

RESULT 1

S05603

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C:Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from P1
A:Reference number: S04850; MUID:89345116; PMID:2668887
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL2>
A:Cross-references: EMBL:X15063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 100.0%; Score 8424; DB 2; Length 1639;

Best Local Similarity 100.0%; Pred. No. 1.5e-253;

Matches 1639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTIFFLCSPFLTIINTQCVTHESYQELVKKLEALEDAVLGTSLFOKEKWLNEGTS 60
Db 1 MKTIFFLCSPFLTIINTQCVTHESYQELVKKLEALEDAVLGTSLFOKEKWLNEGTS 60
QY 61 AVTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGNSRRTNPSDNSS 120
Db 61 AVTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGNSRRTNPSDNSS 120
QY 121 DAKSYADLKHRVRYLLTIKELYPQLFDLTNHLMLTCDNIHGFYKYLIDYEEINELLY 180
Db 121 DAKSYADLKHRVRYLLTIKELYPQLFDLTNHLMLTCDNIHGFYKYLIDYEEINELLY 180
QY 181 LNFYFDLLRAKLDVNCANDYQCPFNKIRANELDVLKLVFGYRPLDNIKDNVGMED 240
Db 181 LNFYFDLLRAKLDVNCANDYQCPFNKIRANELDVLKLVFGYRPLDNIKDNVGMED 240
QY 241 YIKKNKTTIENINELIEESKTTIDKNKNATKEBEKKLYQAOYDLSYNNKOLEEAHNL 300
Db 241 YIKKNKTTIENINELIEESKTTIDKNKNATKEBEKKLYQAOYDLSYNNKOLEEAHNL 300
QY 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANGNTPTNLLDNKKTEEHEKEIKETAKT 360
Db 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANGNTPTNLLDNKKTEEHEKEIKETAKT 360
QY 361 IKENIDSLFDPLELYLREKKNIDISAKVETKESTENEPNGVTYPLSYNDINNAL 420
Db 361 IKENIDSLFDPLELYLREKKNIDISAKVETKESTENEPNGVTYPLSYNDINNAL 420
QY 421 NELNSFGDLINPDYTKEPSKNITYDNERKKFNEIKKIKIEKKKTESDKSYEDRSKS 480
Db 421 NELNSFGDLINPDYTKEPSKNITYDNERKKFNEIKKIKIEKKKTESDKSYEDRSKS 480

```
QY 481 LNDITKEYEKLLEIYDSKFNNNIDLTNFEKMMGKRYSYKVEKLTHHNTFASYENSKHNL 540
Db 481 LNDITKEYEKLLEIYDSKFNNNIDLTNFEKMMGKRYSYKVEKLTHHNTFASYENSKHNL 540
QY 541 EKLTKALKYMEDYSLRNIVVEKELKYKNNLSKIEIENIETLVENIKKDEQOLFEEKITKD 600
Db 541 EKLTKALKYMEDYSLRNIVVEKELKYKNNLSKIEIENIETLVENIKKDEQOLFEEKITKD 600
QY 601 ENKPDEKILEVSDIVKQVQVKVLLMNNKIDELKKTQLILKNVELKHHNIHVPNSYKQENKQ 660
Db 601 ENKPDEKILEVSDIVKQVQVKVLLMNNKIDELKKTQLILKNVELKHHNIHVPNSYKQENKQ 660
QY 661 PYLIVLKEIDKLVKMPKVESLINEEKNNIKTEQSDNSPSTGEITGOATTKPGQO 720
Db 661 PYLIVLKEIDKLVKMPKVESLINEEKNNIKTEQSDNSPSTGEITGOATTKPGQO 720
QY 721 AGSALGDSVQAQAOBQKQAPVPVPPEAKAQVTPPAPVNNKTENYSKLDYLEKLYE 780
Db 721 AGSALGDSVQAQAOBQKQAPVPVPPEAKAQVTPPAPVNNKTENYSKLDYLEKLYE 780
QY 781 FLNTSYICHKYILVSHSTHNEKILQYKITKEESKLSKCDPLDLLFNIONNIPVMSMF 840
Db 781 FLNTSYICHKYILVSHSTHNEKILQYKITKEESKLSKCDPLDLLFNIONNIPVMSMF 840
QY 841 DSLNNSLSOLFMEIYEKEMVCNLYKLDKNDKTKNLLERAKKYVSTSVKTLSSSSMOPLSLT 900
Db 841 DSLNNSLSOLFMEIYEKEMVCNLYKLDKNDKTKNLLERAKKYVSTSVKTLSSSSMOPLSLT 900
QY 901 POKPEVSANDDTSHSTNLNLSLKLFENILSLGKNKNYQELIGOKSSNFYEKILKDS 960
Db 901 POKPEVSANDDTSHSTNLNLSLKLFENILSLGKNKNYQELIGOKSSNFYEKILKDS 960
QY 961 TFYNESFTNFVKSADDDINSANDESKRKKLEEDINKLKTQLQSFPLYKYLKLERLPD 1020
Db 961 TFYNESFTNFVKSADDDINSANDESKRKKLEEDINKLKTQLQSFPLYKYLKLERLPD 1020
QY 1021 KKTGVKYMQJTKLTLLKEQLESKLNSLNNPKHVLQNFVFNKKKEAEIAETENTLEN 1080
Db 1021 KKTGVKYMQJTKLTLLKEQLESKLNSLNNPKHVLQNFVFNKKKEAEIAETENTLEN 1080
QY 1081 TKILLKHVKLYYNGESSPLKTLSESIQTEDNVASLENKVLKSLGKLDNLNLEK 1140
Db 1081 TKILLKHVKLYYNGESSPLKTLSESIQTEDNVASLENKVLKSLGKLDNLNLEK 1140
QY 1141 KKLSSLSSGLHHLIAELKEVIKNNYTGNSPENNTDVNNALLESYKFLPECTDVATVYS 1200
Db 1141 KKLSSLSSGLHHLIAELKEVIKNNYTGNSPENNTDVNNALLESYKFLPECTDVATVYS 1200
QY 1201 ESGSDTLEQSQPKKPASTHVGAESNTITTSQNVDDVDDVIIVPFGSEEDYDDLQGVV 1260
Db 1201 ESGSDTLEQSQPKKPASTHVGAESNTITTSQNVDDVDDVIIVPFGSEEDYDDLQGVV 1260
QY 1261 TGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLKKOLENNVMTFNVNVDIILNSRN 1320
Db 1261 TGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLKKOLENNVMTFNVNVDIILNSRN 1320
QY 1321 KRENFNVLLESOLIPYKDLTSSNVYKDPYKFLNKEKRDKFLSSYNYTKDSIDTDINFAN 1380
Db 1321 KRENFNVLLESOLIPYKDLTSSNVYKDPYKFLNKEKRDKFLSSYNYTKDSIDTDINFAN 1380
QY 1381 DVLGYKYLSEKYSKLDLSIKKYINDKQGENEKYLPFLNNIETLKYTNNDKIDLFVHLE 1440
Db 1381 DVLGYKYLSEKYSKLDLSIKKYINDKQGENEKYLPFLNNIETLKYTNNDKIDLFVHLE 1440
QY 1441 AKVLNYYEKSNNVEYKIKELNLYTKTDKADFKKNNNPFVGLADLSTYNNHNNLTFKFLS 1500
Db 1441 AKVLNYYEKSNNVEYKIKELNLYTKTDKADFKKNNNPFVGLADLSTYNNHNNLTFKFLS 1500
QY 1501 TGMVFNENLAKTVLSNLLDGNLQGMNLNIHQHCVKVKQCPQNSGCFRHLDERECKCLLNYK 1560
Db 1501 TGMVFNENLAKTVLSNLLDGNLQGMNLNIHQHCVKVKQCPQNSGCFRHLDERECKCLLNYK 1560
QY 1561 QBGDKCVENPNPNCNENNGCGDADAKCTEEDSGSGNGKKTCTCKTDPDYPFDGIFCSCSS 1620
```

```
Db 1561 QBGDKCVENPNPNCNENNGCGDADAKCTEEDSGSGNGKKTCTCKTDPDYPFDGIFCSCSS 1620
QY 1621 NFLGISFLLILMLILYSFI 1639
Db 1621 NFLGISFLLILMLILYSFI 1639

RESULT 2
A:24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodi
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C:Accession: A24594
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholl
Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of P1
A:Reference number: A24594; MUID:86014355; PMID:2995820
A:Accession: A24594
A:Molecule type: DNA
A:Residues: 1-1640 <HOL>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 97.2%; Score 8191.5; DB 2; Length 1640;
Best Local Similarity 97.9%; Pred. No. 2.3e-246;
Matches 1605; Conservative 8; Mismatches 26; Indels 1; Gaps 1;

QY 1 MKIIIFLCFLFFIINTQCVTTHESYQELVKLEALEDAVLTSYLFQKQKMWLNCEGTSGT 60
Db 1 MKIIIFLCFLFFIINTQCVTTHESYQELVKLEALEDAVLTSYLFQKQKMWLNCEGTSGT 60
QY 61 AVTTSTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVS 120
Db 61 AVTTSTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVS 120
QY 121 DAKSYADLKHVRNVLTLTKELKYPQLFDTLNHMLTLCNDNIHGFYKYLIDGYEBEINELLYK 180
Db 121 DAKSYADLKHVRNVLTLTKELKYPQLFDTLNHMLTLCNDNIHGFYKYLIDGYEBEINELLYK 180
QY 181 LNFYDILLRAKLVNDVANDYCOIPNLKIRANELDVLKLVFGYRKLPLNDIKDNVCKMED 240
Db 181 LNFYDILLRAKLVNDVANDYCOIPNLKIRANELDVLKLVFGYRKLPLNDIKDNVCKMED 240
QY 241 YIKKNNKTITENIELIEESKKTIDKNKNATKEEEKKLYQAOYDLSIYNKQLEEAHNLIS 300
Db 241 YIKKNNKTITENIELIEESKKTIDKNKNATKEEEKKLYQAOYDLSIYNKQLEEAHNLIS 300
QY 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANGSNTPNTLLDKNNKIEHEKEIKEIAKT 360
Db 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANGSNTPNTLLDKNNKIEHEKEIKEIAKT 360
QY 361 IKFNIDSLFTDPLEYLYLREKNNKIDISAKVETKESTEPNPGVTYPLSYNDINNAL 420
Db 361 IKFNIDSLFTDPLEYLYLREKNNKIDISAKVETKESTEPNPGVTYPLSYNDINNAL 420
QY 421 NELNSFGDLINFPDYTKEPSKNIYTDNERKKFINEIKERKIEKKIESDKKSYEDRSKS 480
Db 421 NELNSFGDLINFPDYTKEPSKNIYTDNERKKFINEIKERKIEKKIESDKKSYEDRSKS 480
QY 481 LNDITKEYEKLLEIYDSKFNNNIDLTNFEKMMGKRYSYKVEKLTHHNTFASYENSKHNL 540
Db 481 LNDITKEYEKLLEIYDSKFNNNIDLTNFEKMMGKRYSYKVEKLTHHNTFASYENSKHNL 540
QY 541 EKLTKALKYMEDYSLRNIVVEKELKYKNNLSKIEIENIETLVENIKKDEQOLFEEKITKD 600
Db 541 EKLTKALKYMEDYSLRNIVVEKELKYKNNLSKIEIENIETLVENIKKDEQOLFEEKITKD 600
QY 601 ENKPDEKILEVSDIVKQVQVKVLLMNNKIDELKKTQLILKNVELKHHNIHVPNSYKQENKQ 660
Db 601 ENKPDEKILEVSDIVKQVQVKVLLMNNKIDELKKTQLILKNVELKHHNIHVPNSYKQENKQ 660
QY 661 PYLIVLKEIDKLVKMPKVESLINEEKNNIKTEQSDNSPSTGEITGOATTKPGQO 720
Db 661 PYLIVLKEIDKLVKMPKVESLINEEKNNIKTEQSDNSPSTGEITGOATTKPGQO 720
QY 721 AGSALGDSVQAQAOBQKQAPVPVPPEAKAQVTPPAPVNNKTENYSKLDYLEKLYE 780
Db 721 AGSALGDSVQAQAOBQKQAPVPVPPEAKAQVTPPAPVNNKTENYSKLDYLEKLYE 780
QY 781 FLNTSYICHKYILVSHSTHNEKILQYKITKEESKLSKCDPLDLLFNIONNIPVMSMF 840
Db 781 FLNTSYICHKYILVSHSTHNEKILQYKITKEESKLSKCDPLDLLFNIONNIPVMSMF 840
QY 841 DSLNNSLSOLFMEIYEKEMVCNLYKLDKNDKTKNLLERAKKYVSTSVKTLSSSSMOPLSLT 900
Db 841 DSLNNSLSOLFMEIYEKEMVCNLYKLDKNDKTKNLLERAKKYVSTSVKTLSSSSMOPLSLT 900
QY 901 POKPEVSANDDTSHSTNLNLSLKLFENILSLGKNKNYQELIGOKSSNFYEKILKDS 960
Db 901 POKPEVSANDDTSHSTNLNLSLKLFENILSLGKNKNYQELIGOKSSNFYEKILKDS 960
QY 961 TFYNESFTNFVKSADDDINSANDESKRKKLEEDINKLKTQLQSFPLYKYLKLERLPD 1020
Db 961 TFYNESFTNFVKSADDDINSANDESKRKKLEEDINKLKTQLQSFPLYKYLKLERLPD 1020
QY 1021 KKTGVKYMQJTKLTLLKEQLESKLNSLNNPKHVLQNFVFNKKKEAEIAETENTLEN 1080
Db 1021 KKTGVKYMQJTKLTLLKEQLESKLNSLNNPKHVLQNFVFNKKKEAEIAETENTLEN 1080
QY 1081 TKILLKHVKLYYNGESSPLKTLSESIQTEDNVASLENKVLKSLGKLDNLNLEK 1140
Db 1081 TKILLKHVKLYYNGESSPLKTLSESIQTEDNVASLENKVLKSLGKLDNLNLEK 1140
QY 1141 KKLSSLSSGLHHLIAELKEVIKNNYTGNSPENNTDVNNALLESYKFLPECTDVATVYS 1200
Db 1141 KKLSSLSSGLHHLIAELKEVIKNNYTGNSPENNTDVNNALLESYKFLPECTDVATVYS 1200
QY 1201 ESGSDTLEQSQPKKPASTHVGAESNTITTSQNVDDVDDVIIVPFGSEEDYDDLQGVV 1260
Db 1201 ESGSDTLEQSQPKKPASTHVGAESNTITTSQNVDDVDDVIIVPFGSEEDYDDLQGVV 1260
QY 1261 TGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLKKOLENNVMTFNVNVDIILNSRN 1320
Db 1261 TGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLKKOLENNVMTFNVNVDIILNSRN 1320
QY 1321 KRENFNVLLESOLIPYKDLTSSNVYKDPYKFLNKEKRDKFLSSYNYTKDSIDTDINFAN 1380
Db 1321 KRENFNVLLESOLIPYKDLTSSNVYKDPYKFLNKEKRDKFLSSYNYTKDSIDTDINFAN 1380
QY 1381 DVLGYKYLSEKYSKLDLSIKKYINDKQGENEKYLPFLNNIETLKYTNNDKIDLFVHLE 1440
Db 1381 DVLGYKYLSEKYSKLDLSIKKYINDKQGENEKYLPFLNNIETLKYTNNDKIDLFVHLE 1440
QY 1441 AKVLNYYEKSNNVEYKIKELNLYTKTDKADFKKNNNPFVGLADLSTYNNHNNLTFKFLS 1500
Db 1441 AKVLNYYEKSNNVEYKIKELNLYTKTDKADFKKNNNPFVGLADLSTYNNHNNLTFKFLS 1500
QY 1501 TGMVFNENLAKTVLSNLLDGNLQGMNLNIHQHCVKVKQCPQNSGCFRHLDERECKCLLNYK 1560
Db 1501 TGMVFNENLAKTVLSNLLDGNLQGMNLNIHQHCVKVKQCPQNSGCFRHLDERECKCLLNYK 1560
QY 1561 QBGDKCVENPNPNCNENNGCGDADAKCTEEDSGSGNGKKTCTCKTDPDYPFDGIFCSCSS 1620
```

Db 661 PYLIVLKKKEIDKLKVFMPKVESLINEEKKNIKTQGSNDSEPTGEITGQATTPGQ 720
QY 721 AGSALEGDSVQAQAQSQKQAQPPVPPVPAKAQVTPPPAPVNNKTEVNSKLDYLEKLYE 780
Db 722 AGSALEGDSVQAQAQSQKQAQPPVPPVPAKAQVTPPPAPVNNKTEVNSKLDYLEKLYE 780
QY 781 FLNTSYICHKYILVSHSTWNEKILKYKITKEESKLSSCDPLDLFNQNNIPVMSMF 840
Db 781 FLNTSYICHKYILVSHSTWNEKILKYKITKEESKLSSCDPLDLFNQNNIPVMSMF 840
QY 841 DSLNLSQLFMEIYKEMVCNLYKLDNDKIKNLEEAQKVTSTVKTLSSSSQPLSLT 900
Db 841 DSLNLSQLFMEIYKEMVCNLYKLDNDKIKNLEEAQKVTSTVKTLSSSSQPLSLT 900
QY 901 PQDKPEVSANDTSHSTNLSNLSKLFENILSLGKNKIYQELIGOKSSSENFYKILKDS 960
Db 901 PQDKPEVSANDTSHSTNLSNLSKLFENILSLGKNKIYQELIGOKSSSENFYKILKDS 960
QY 961 TFYNESFTNFVKSADDSLNDLNSDKRKKLEEDINKLKTQLSFDLYNKYKILKLERLFD 1020
Db 961 TFYNESFTNFVKSADDSLNDLNSDKRKKLEEDINKLKTQLSFDLYNKYKILKLERLFD 1020
QY 1021 KKTIVGKYKMOIKKLLKLEQESKLSNPNKPHVLQNSVFNKKKEAEIAETENTLEN 1080
Db 1021 KKTIVGKYKMOIKKLLKLEQESKLSNPNKPHVLQNSVFNKKKEAEIAETENTLEN 1080
QY 1081 TKILLKHYGLVYKNGESSPLKTLSEESTQTEDNEDNFKVLSKLEKLDNLEK 1140
Db 1081 TKILLKHYGLVYKNGESSPLKTLSEESTQTEDNEDNFKVLSKLEKLDNLEK 1140
QY 1141 KKLVSLSGLLHLLIAELKEVINKNKTGNSPSENNTDVNNALSYKFFPEGTDVATVVS 1200
Db 1141 KKLVSLSGLLHLLIAELKEVINKNKTGNSPSENNTDVNNALSYKFFPEGTDVATVVS 1200
QY 1201 EGSQDLEQSQKPPASTHVGAESNITTSQNVDDDEVDDVILVIFGESEEDYDGLGOV 1260
Db 1201 EGSQDLEQSQKPPASTHVGAESNITTSQNVDDDEVDDVILVIFGESEEDYDGLGOV 1260
QY 1261 TGEAVTSPVIDNLSKIENEYEVLYLKLPLAGVYRSLSKQLENNVMTFNVNVDILNSRFN 1320
Db 1261 TGEAVTSPVIDNLSKIENEYEVLYLKLPLAGVYRSLSKQLENNVMTFNVNVDILNSRFN 1320
QY 1321 KRFNKNVLESLLPYKDTSSVYVVDKPYKFLNKEKROKFLSSYNIKDSITDINFAN 1380
Db 1321 KRFNKNVLESLLPYKDTSSVYVVDKPYKFLNKEKROKFLSSYNIKDSITDINFAN 1380
QY 1381 DVLGYKILSEKYKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFIHLE 1440
Db 1381 DVLGYKILSEKYKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFIHLE 1440
QY 1441 AKVLNYYEKSNNVEVKIKELNYLKTQDLADFKKNNFVGADLSTQYNNHLLTKFLS 1500
Db 1441 AKVLNYYEKSNNVEVKIKELNYLKTQDLADFKKNNFVGADLSTQYNNHLLTKFLS 1500
QY 1501 TGMVFENLAKTVLSNLDGNL-OGMLNIHQCVKQKQPCQNSGCFRHLDERECKLLNY 1560
Db 1501 TGMVFENLAKTVLSNLDGNL-OGMLNIHQCVKQKQPCQNSGCFRHLDERECKLLNY 1560
QY 1560 KQSGDKCVENPNTCNENGGCCADAKCTEEDSGSGNGKKTCTCTKPDSPYPLDGFICSS 1619
Db 1561 KQSGDKCVENPNTCNENGGCCADAKCTEEDSGSGNGKKTCTCTKPDSPYPLDGFICSS 1620
QY 1620 SNFLGIFSLLILMLIYSFI 1639
Db 1621 SNFLGIFSLLILMLIYSFI 1640

RESULT 3
SAZOK1

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st

C:Species: Plasmodium falciparum

C:date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000

C:Accession: A25120
R:MacKay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.
EMBO J. 4, 3823-3829, 1985
A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f
A:Reference number: A91030; MUID:86136024; PMID:3004972
A:Accession: A25120
A:Molecule type: DNA
A:Residues: 1-1631 <MAC>
C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane pr
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F:67-84/Region: 3-residue repeats (S-G-T/P)
F:1614-1631/Domain: membrane anchor #status predicted <MBN>
F:97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn

Query Match 94.6%; Score 7965; DB 1; Length 1631;
Best Local Similarity 95.3%; Pred. No. 2.4e-239;
Matches 1563; Conservative 19; Mismatches 48; Indels 10; Gaps 3;

QY 1 MKTIIFLCFLPFIINTQCVTHESYOELVKKLEALEDAVLTYGSLFQKQKMWLNEGTSGT 60
Db 1 MKTIIFLCFLPFIINTQCVTHESYOELVKKLEALEDAVLTYGSLFQKQKMWLNEBEITT 60
QY 61 AVTSTPGSGSVASGGSGSVASGGSVASGGSVASGGSGSNRRTPNSDSSDS 120
Db 61 KGASQSGTSGT--SGTSGPSGSGTSPSSRSNTLPRSNITSSGAS-----PPADASDS 111
QY 121 DAKSYADLKHVRVNYLLTTKELYPQLFDLTNHLMTLCNIGHGFKYLDIGYEBINELLYK 180
Db 112 DAKSYADLKHVRVNYLLTTKELYPQLFDLTNHLMTLCNIGHGFKYLDIGYEBINELLYK 171
QY 181 LNFYFLLRAKLDNCVANDYQCPENLKIRANELDVLKLVFGYRPLDNIKDNVGMKED 240
Db 172 LNFYFLLRAKLDNCVANDYQCPENLKIRANELDVLKLVFGYRPLDNIKDNVGMKED 231
QY 241 YIKNNKTTIENINELIEESKTTIDKNKATKEEKKKLYQAQYDLSIYNKQLEEAHNLIS 300
Db 232 YIKNNKTTIENINELIEESKTTIDKNKATKEEKKKLYQAQYDLSIYNKQLEEAHNLIS 291
QY 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNLLDKNKKIEEHEKEIKETAKT 360
Db 292 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNLLDKNKKIEEHEKEIKETAKT 351
QY 361 IKFNIDSLTDPLELYVLRKKNKIDISAKVETKESTENEPNGVYTPLSYNDNNAL 420
Db 352 IKFNIDSLTDPLELYVLRKKNKIDISAKVETKESTENEPNGVYTPLSYNDNNAL 411
QY 421 NELNSFGDLINPDYTKEPSKNIYTDNERKKFTNEIKEIKIEKKKIESDKKSYEDRSKS 480
Db 412 NELNSFGDLINPDYTKEPSKNIYTDNERKKFTNEIKEIKIEKKKIESDKKSYEDRSKS 471
QY 481 LNDITKEYEKLLENIYDSKFNNNIDLTNFEKMMGKRYSYKVEKLTHTHNTFASYSKHNL 540
Db 472 LNDITKEYEKLLENIYDSKFNNNIDLTNFEKMMGKRYSYKVEKLTHTHNTFASYSKHNL 531
QY 541 EKLTKALKYMEDYSLRNIYVVEKELKYKNLISKIENEIETLVENIKKDEQLPEFKITKD 600
Db 532 EKLTKALKYMEDYSLRNIYVVEKELKYKNLISKIENEIETLVENIKKDEQLPEFKITKD 591
QY 601 ENKPDEKILVSDIVKQVQKVLMMNKIDELKKTQILKKNVLEKHNTHVPSNYKQENKQE 660
Db 592 ENKPDEKILVSDIVKQVQKVLMMNKIDELKKTQILKKNVLEKHNTHVPSNYKQENKQE 651
QY 661 PYLIVLKKKEIDKLKVFMPKVESLINEEKKNIKTQGSNDSEPTGEITGQATTPGQ 720
Db 652 PYLIVLKKKEIDKLKVFMPKVESLINEEKKNIKTQGSNDSEPTGEITGQATTPGQ 711
QY 721 AGSALEGDSVQAQAQSQKQAQPPVPPVPAKAQVTPPPAPVNNKTEVNSKLDYLEKLYE 780
Db 712 AGSALEGDSVQAQAQSQKQAQPPVPPVPAKAQVTPPPAPVNNKTEVNSKLDYLEKLYE 771

Db 710 IPKVDMKKKEQAVLSS-----ITQPLVAASETTEDGGHSTHTLSQSGETEVTET 760
QY 738 KOAOPV-----PVPVPEAKAQPPTPPAPVNNKTNVSKLDYLEKLYEFLNTSYICH 789
Db 761 ETEETVGHVTTVTITLPKPEKVVVNSTEHSKSDNSQALTTVTYKLDLEFLTYSYICH 820
QY 790 KYILVSHSTMNKILKYKITKEESKLSGCDPLDLLFNFIQNNIPVYSMFDSLANSLSQ 849
Db 821 KYILVSNSSMDQLLEVYNTPEENELKSCDPLDLLFNFIQNNIPAMYSLYDSMNDLQH 880
QY 850 LFWEIYEKEMVCLYKLDNDKIKNLEBAKV-----STSVKTLSSS----- 893
Db 881 LFELYQKEMYYLHKLEENHKKLEQKQITCTSTSSPCNTVTNTAQATHNSON 940
QY 894 --MQPLSTPPOKPEVSAND--DTSHS-----TNLNSLKLFPENILSLGKNKIYOEL-I 943
Db 941 QOSNASSTWQNGVAVSSGPAVVEESHDPPLTVLSTISNDLKGIVSLNLGNKTKVNPPLTI 1000
QY 944 GOKSENFEYKILKSDTPTYNESFTNFVYSKADDDINSLNDESKRKKLEEDINKLKTQL 1003
Db 1001 STTEMEKFENTLKNNDTVFNDIDQFVKSNSKVITGLT-ETQKNALENDEIKKLDLQL 1059
QY 1004 SFDLYNKYKLERLFDKKTGCKYKMOIKLTLLEQLESKLSLNNPKHYLQNFVSFF 1063
Db 1060 SFDLYNKYKLDRLFNKKELGQDKMOIKLTLLEQLESKLSLNNPHNYLQNFVSFF 1119
QY 1064 NKKKEAEIAETENTLTKILLKHYKGLVYNGESSPLKTLSESIQEDNYASLENPK 1123
Db 1120 NKKKEAEIAETENTLTKILLKHYKGLVYNGESSPLKTLSESIQEDNYANLEKER 1179
QY 1124 VLSKLEGLKUNLEKKLSYLSGLHLIAELKEVINKNKYTGNSPENTDNNNALE 1183
Db 1180 VLSKIDGKLDNLHLGKLSLSSGLHQLITELKEVINKNKYTGNSPENKNKYNEALK 1239
QY 1184 SYKKFLPECTDVATVYSE-----SGSDTLQSQPKKPASTHVGAESNTIIT 1229
Db 1240 STENLPE-AKTVTVVPPQDVTSPPLSVRVSGSGSTKEETQIPTSGSLTELOQVVQ 1298
QY 1230 SONVDEVDVIVIPFGESEDDYDQGVVTEAVTPSVIDNLSKIENEYVLYLKLPL 1289
Db 1299 LQNYDEEDSLVPLIFGESEDNDEYDQGVVTEAISVT-MDNILSGFENEYDVLYLKLPL 1357
QY 1290 AGVYRSKLQLENVMTFNVNVDILNSFRNKRNFKNVLESOLIPYKDLTSSNVVVKDP 1349
Db 1358 AGVYRSKLQIEKNFTFNLNLDILNSRLKRRKYFLDLESOLMQFKHISSEYIIDS 1417
QY 1350 YKFLNKRKDRKFLSVNYIKSDIDTDFNANDVLGYKILSKYKSDLSIKKYI----- 1404
Db 1418 FKLLSEQNTLLSKYKIKESVENDIKPAQEGISYEVKLAKYKODLESIRKVIKEKE 1477
QY 1405 -----NDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVLYNTYVE 1449
Db 1478 KFPSSPPTTTPSPVKTDQKESKFLPLFTNLTETLYNVLNKNDDYLLINKAKINDCNVE 1537
QY 1450 KSNVEVKIKELNYLKTIOQKLADFKNNFVGIADLSTDYNNHLLTKPLTSGMVFENIA 1509
Db 1538 KDEAHVYKTKLDLKAIDDKIDLFKNHNDFEAIKLLINDTDKMLGKLLSGLV-QNPP 1596
QY 1510 KTVLSNLLDGNLQGLMNLISQHCVKQKQPCQNSGCFRHLDERECCLLNYKQEGDKCVEN 1569
Db 1597 NTIISKLIEGKFQDMLNISOHQCVKQKQPCQNSGCFRHLDERECCLLNYKQEGDKCVEN 1656
QY 1570 PNPTCNENGGDAADAKCTEEDSGSGKKTCTCTKPDSPYPLFDGIFCFSNFPGLTSPFL 1629
Db 1657 PNPTCNENGGDAADAKCTEEDSGSGKKTCTCTKPDSPYPLFDGIFCFSNFPGLTSPFL 1716
QY 1630 ILMILYISFI 1639
Db 1717 ILMILYISFI 1726

RESULT 6
A54498

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V
Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmo
A:Reference number: A54498; MUID:88142999; PMID:2449612
A:Accession: A54498
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PET>
A:Cross-references: GB:M19143; NID:gl60412; PIDN:AAA29653.1; PID:gl60413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 60.1%; Score 5065; DB 2; Length 1701;
Best Local Similarity 60.1%; Pred. No. 1.3e-149;
Matches 1042; Conservative 227; Mismatches 338; Indels 126; Gaps 29;

QY 1 MKIIIFLCSFLFIINTQCVTHESYOELVKLEALEDAVLTVGYSLFQKMKVNLNCTSGT 60
Db 1 MKIIIFLCSFLFIINTQCVTHESYOELVKLEALEDAVLTVGYSLFQKMKVNLNCTSGT 60
QY 61 AVTTSTPGSKGVSAGSGSGSVASGSGSVASGSGSVASGSGSGSNRRTNPDSNDS 120
Db 61 AVTTSTPGSGSVT---SGGSVASVASVAGG---SGGSVASGSGSGSNRRTNPDSNDS 114
QY 121 DAKSVADLKHVRNVLITIKELYPQLFDLTNHLMTLCDNIHGFKYLIDGYEINELLVK 180
Db 115 NTKTYADLKHVRVNYLFTIKELYPQLFDLTNHLMTLSKNVDGFKYLIDGYEINELLVK 174
QY 181 LNFYDLLRAKLVNDVANDYCOIPFNKLRANELDVLKLVGYRKLPLNIDKNVGMED 240
Db 175 LNFYDLLRAKLVNDVANDYCOIPFNKLRANELDVLKLVGYRKLPLNIDKNVGMED 234
QY 241 YTKKNTKTIENTNELIEESKKTIDKNKNATKEEKKLYQAOYDYSIYNKQLEEAHNLIS 300
Db 235 YTKKNTKTIENTNELIEESKKTIDQKNADNEEGKKLYQAOYVNFYINQKQLEEAHNLIS 294
QY 301 VLEKRTDITLKKNENIKELDKINEIK---NPPANSNGTNPNTLLD-KNKKTEEHEKEIK 355
Db 295 VLEKRTDITLKKNENIKELDKINEIK---TTGSKPNPLPENKKEVEGHEEKIK 351
QY 356 EIAKTIKFNIDSLFDPLELYYLRKKNKIDISAKVE--TKRESTPN-EYPNGVYTPLS 412
Db 352 EIAKTIKFNIDSLFDPLELYYLRKKNKVDVTPKSQDTPKSVQIPKVPYNGIYVPLP 411
QY 413 YNDINNAL---NELNSFGDLINPFDYTKPSKNIYTDN-ERKKFNIKEIKIEKKKI- 467
Db 412 LTDIHNLSAADNDKNSYGDLMNP-DTKEKINEKIITDNKERKIFINNIKKQIDLEEKNIN 470
QY 468 ---ESDKSYEDRSKSLNDITKEYEKLLNEIYDSFNNDILTNFEKMMGKRYSYKVEKL 524
Db 471 HTEKONKKLLEDEYKS---KKDYELLEKEFYEMFNNFNDKDDVDKIFARSITYNVEKQ 526
QY 525 THHNTFASYSKHNKLEKTKALKYMEDYSLRNIVVEKELKYKNLISKIEIETLVEN 584
Db 527 RYNNKFSSNNSVYVQKLLKALSLEYDSLGRKISEKDNHYTTLKTGLEADIKKLTTEE 586
QY 585 IKKDEQLFEKK---ITKDNKPDEKILEVSDIVKVVQVQVLLMKNKIDELKKTQLILKNV 641
Db 587 IKSSENKILEKNFKGLTHSANAS---LEVSDIVKLQVQVLLIKKIEDLRKLTFLKNA 642
QY 642 ELKHNHVPNSYKQENKQEPYLYLVLLKEDIKLVKMPKVESLINEEKNKITEGSDNS 701
Db 643 QLKDSIHVPNIYKPNKPEPYLYLVLLKEDIKLVKMPKVESLINEEKNKITEGSDNS 695
QY 702 EPSTEGETITQATTKPGQAGSALGSDSVQAQAEQKQA-----QPPVPVPEAKAQ 754
Db 696 --ITQPLVAASETTEDGGHSTHTL-SQSGETEVTETVGHVTTVTITLPKEES 752
QY 755 VPTPPAPVNNKTN-----VSKLDYLEKLYEFLNTSYICHYILVSHSTMNEKILQ 806


```
Db 753 APREVKVVENSIEHKSNDNSQALTKTVYKKLDEFLLTKSYICHKYILVSNSSMDQKLEV 812
QY 807 YKLTKEESKSCDPLDLLFNIONNIPVMSFDSLNNSLSOLFWEIYEKEMVCNLYKL 866
Db 813 YNLTPEENELKSCDPLDLLFNIONNIPVMSFDSLNNSLSOLFWEIYEKEMVCNLYKL 872
QY 867 KMDKIKNLEAKV-----STSVKTLSSSS-----MQLSLTPQDKPEVS 908
Db 873 KEENHIKLLLEQKQITGTSSTSPGNTVNTAQSAATHSNQOQSNASTNTQNGVAVS 932
QY 909 AND---DTSHS-----TNLNNSLKFNILSLGKKNKIYQEL-IGQKSSNFYKILKDS 960
Db 933 SGPAVVEESHDPVLISNDLKGIVSLNNGKTKVPNPLTISTTEMEKFYENILKND 992
QY 961 TFYNESFTNPKSKADDIINSNDKSRKLEEDINKLKTLOISFDLYNKYKLEERLEF 1020
Db 993 TYFNDDIKOFVNSKVITGLT-ETOKNALNDEIKLKDTLOISFDLYNKYKLEERLEF 1051
QY 1021 KKTGVGYKMOIKKLLKLEQLESKLSLNNPKHVLQNFVFFNKKKEAEIAETENTLEN 1080
Db 1052 KKLELQDKMOIKKLLKLEQLESKLSLNNPHVNLQNFVFFNKKKEAEIAETENTLEN 1111
QY 1081 TKILLHYKGLVYNGESSPLKTLSEESTQTEDNYASUENFKVLSKLEGLKDNLENLEK 1140
Db 1112 TKILLHYKGLVYNGESSPLKTLSEVSTQTEDNYANLEKPRALSCKIDGKLDNLEHLG 1171
QY 1141 KLSYLSGGLHLLIAELKEVINKNKTGNSPSENNTDVNNALESYKFLPEGTDVATVVS 1200
Db 1172 KLSYLSGGLHLLIAELKEVINKNKTGNSPSENNTDVNNALESYKFLPEGTDVATVVS 1230
QY 1201 E-----SGSDTLEQSPKPAKPAHVGASNTITTSQNVDDVDVIVTPIF 1246
Db 1231 PPQPDVTPSLSVRVSSTGKEEQIPTSGLLTELQVQVQVQVDEEDSLVLP 1290
QY 1247 GESEEDYDLQGVVTEGATPSVIDNLSKIENEYEVYIKPLAGVYRSKUKOLENNVT 1306
Db 1291 GESEEDYDLQGVVTEGATPSVIDNLSKIENEYEVYIKPLAGVYRSKUKOLENNVT 1349
QY 1307 FNVNVDILNSRKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNF 1366
Db 1350 FNVNVDILNSRKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNFKNF 1409
QY 1367 YIKSDITDINFANDVGYKILSEYKSDLSIKYI-----ND 1406
Db 1410 YIKESVENDIKPAQEGISYKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTD 1469
QY 1407 KQGENEKYLPFLNNIETLYTVNDKIDLVHLEAKVNLVTEKSNVEYKIKELNLYKTI 1466
Db 1470 EQKESKFLPFLNIETLYTVNDKIDLVHLEAKVNLVTEKSNVEYKIKELNLYKTI 1529
QY 1467 QDKLADFPKNNFVGIADLSTDVNHNLLTKFLSTGMVFENLAKTVLSNLLDGLQGLN 1526
Db 1530 DDKLADFPKNNFVGIADLSTDVNHNLLTKFLSTGMVFENLAKTVLSNLLDGLQGLN 1588
QY 1527 ISOHQVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNTCENNINGGCDADAK 1586
Db 1589 ISOHQVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNTCENNINGGCDADAT 1648
QY 1587 CTEEDSGNCKKTKCTCTKPDSPYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1639
Db 1649 CTEEDSGNCKKTKCTCTKPDSPYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701
```

RESULT 7

A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C:Species: Plasmodium falciparum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C:Accession: A26868
R:Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium

A:Reference number: A26868; MUID:88011243; PMID:3079521
A:Accession: A26868
A:Molecule type: DNA
A:Residues: 1-1701 <TAN>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 60.0%; Score 5053; DB 2; Length 1701;
Best Local Similarity 60.1%; Pred. No. 2.9e-149;
Matches 1042; Conservative 225; Mismatches 338; Indels 130; Gaps 30;
QY 1 MKIIFLCFLFIINTQCVTHESYQELVKLEDAVLGTYSLFQKEMVNEGTS 60
Db 1 MKIIFLCFLFIINTQCVTHESYQELVKLEDAVLGTYSLFQKEMVNEGTS 60
QY 61 AVTSTPGSGSVASGGSGSVASGGSGSVASGGSGSVASGGSGSRRTPNSDSS 120
... Db 61 AVTSTPGSGSVT---SGGSVASVASVSG---SGGSVASGGSGSRRTPNSDSS 114
QY 121 DAKSYADLKHRYNLLTIKELKYPOLFDTNHLMTLCNIGHGFKYLDIGEYINELLYK 180
Db 115 NTQYADLKHRYNLLTIKELKYPOLFDTNHLMTLSKNVDGFKYLDIGEYINELLYK 174
QY 181 LNFYDILLRAKLDVANDVCQIPFNLKIRANELDVLKLVFGYRKPDLNIDKONVGKMD 240
Db 175 LNFYDILLRAKLDVANDVCQIPFNLKIRANELDVLKLVFGYRKPDLNIDKONVGKMD 234
QY 241 YTKNKKTTININELIEESKTTIDKNKNTKBEKKKLYQAOYDLYNNKQLEAHNLIS 300
Db 235 YTKNKKTTIANINELIEGSKTTIDQKNADNEEGKKLYQAOYNLFYNNKQLEAHNLIS 294
QY 301 VLEKRTDTLKKENIKELDKINEIK---NPPANGNTPTNLLD-KNKKIEEHEKEIK 355
Db 295 VLEKRTDTLKKENIKELDKINEIK---NPPANGNTPTNLLD-KNKKIEEHEKEIK 351
QY 356 EIATIKFIDSLFTDPLEYLYREKKNINIDISAKVE--TKRESTEPN-EYPNGVYPLS 412
Db 352 EIATIKFIDSLFTDPLEYLYREKKNKVDVTPKSQDPTKSVQIPKVPYNGVYPLP 411
QY 413 YNDINNAL---NELNSFGDLINFPDYTKPSKIYTDN-ERKAFINEIKIEKKKI- 467
Db 412 LTDIHSNLAADNKNISYGLMNP-DTKEKINEXIITDNKERKIFINNIKKQIDLEKNIN 470
QY 468 ---ESDKSYEDRSKSLNDITREYKLLNEIYDSKNNINIDLTFEKKMMKRYSYVEKL 524
Db 471 HTKEQNKLLLEDYKS---KKDYELLEKFEYEMKNNFDDKVDVKIFISARTYVNEKQ 526
QY 525 THHTNFASYENSKHNLEKTKALKYMEDYSLRNIVVEKELKYKNIISKIENIEIETLVEN 584
Db 527 RYNNKFSNNNSVYVQKLSYLEDYSLRKGISEKDFNHYYTLTKTGLADIKKLTTEE 586
QY 585 IKKDEQLPEKK---ITKQENKPEKILEVSDIVKQVQVQVLLMKNIDELKTKQLILKNV 641
Db 587 IKSSNKKILEKPKGITHSANAS-----LEVSQIVKQVQVLLIKKIEDLURKLEFLKNA 642
QY 642 ELKHNTHVNSPKQENKQPPYLLIVLKEIDKLVFMPKVESLINEKKNIKTEGQSDNS 701
Db 643 QLKDSLHVNPNIYKPNKPEPYLLIVLKEVDKLEKIPKVKMDLKKKEQAVLSS----- 695
QY 702 EPSTEGETQATTKPGQAGSALGDSVQAOAQEQKA-----OPPVVPVPEAKAQ 754
Db 696 --ITQPLVAASETTEDGGHSTHTL-SQSGETEVEETEETVGHVTTTITLPPKEES 752
QY 755 VTPPPAPVNNKTEN-----VSKLDYLEKLYEFLNTSYICHKYILVSHSTHNEKLLKQ 806
Db 753 APKEVAVVNSIEHKGNDNSQALTKTVYKKLDEFLLTKSYICHKYILVSNSSMDQKLEV 812
QY 807 YKLTKEESKSCDPLDLLFNIONNIPVMSFDSLNNSLSOLFWEIYEKEMVCNLYKL 866
Db 813 YNLTPEENELKSCDPLDLLFNIONNIPVMSFDSLNNSLSOLFWEIYEKEMVCNLYKL 872

Qy	1420	N I E T L Y T V N D K I D L F V I H L E A K V L N Y T Y E K S N E V E V K I E L N Y L K T I O D K L A D F K K N N F	1479
Db	1517	S L O Q E Y E S L S K V N T Y T D N L K K V I N N C K L E K T E A E T Y T K K I Q D Y N K M D E K L E E Y K K	1572
Qy	1480	V G I A D L S T D Y N H N N L L T K F L S T C M W F E N L A K T V L S N L L D G N L O G M L N I S Q H O C V K K O C P Q	1539
Db	1573	---S E K N E V K S G L L E K L M K S K L I K E N E S K E I L S O L L N V Q T L T M S S E H T C I D T N V P D	1629
Qy	1540	N S G C F R H L D R E E C K L L N Y K Q E D C K V E N P N P T C N E N N G G S C D A D A K T E E D S G N K K I	1599
Db	1630	N A A C Y R Y L D G W E R C L L F E K E G G K C V P S N V T C K D N N G C A P E A E K W M T D S ---N K I	1685
Qy	1600	T C E C T K P D S V L P D G I F C S S N F L G I S F L L I M L I L	1635
Db	1686	V C K T K E G S E P L F E G V F C S S S F L S F L L M L L F L	1721
RESULT 11			
S06286			
Major merozoite surface antigen precursor - malaria parasite (Plasmodium fa			
N;Alternate names: 190K protein; polymorphic schizont antigen			
C;Species: Plasmodium falciparum			
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jun-2000			
C;Accession: S06286			
R;Certa. U.; Rotmann, D.; Matile, H.; Reber-Liske, R.			
EMBO J. 6, 4137-4142, 1987			
A;Title: A naturally occurring gene encoding the major surface antigen prec			
A;Reference number: S06286; MUID:88166657; PMID:3327688			
A;Accession: S06286			
A;Status: not compared with conceptual translation			
A;Molecule type: DNA			
A;Residues: 1-1060 <CER>			
C;Superfamily: major merozoite surface antigen			
C;Keywords: surface antigen			
Query Match 33.6%; Score 2829.5; DB 2; Length 1060;			
Best Local Similarity 55.5%; Pred. No. 1e-80;			
Matches 609; Conservative 145; Mismatches 255; Indels 89; Gaps			
Qy	1	M K I I F F C S L F F I I N T Q C V T H E S Y Q E L V K K L E A L E D A V L T G Y S L F Q E K M V L N E G T S G T	60
Db	1	M K I I F F C S L F F I I N T Q C V T H E S Y Q E L V K K L E A L E D A V L T G Y S L F Q E K M V L D G A N T Q	60
Qy	61	A V T T S T P G S K G S V A S G G S G S V A S G S V A S G S G S G S G S R R T N P S D N S S D S	120
Db	61	V W A K P A D A V S T Q A K N P P G A T V P S -----C T A S T K G A I R S P G A -----A N P S D D S S D S	108
Qy	121	D A K S Y A D L K H R V N Y L L T K E L Y P O L F D L T N H M L T L C D N I H G F K Y L I D G Y E E I N E L L Y K	180
Db	109	D A K S Y A D L K H R V Q N Y L F T I K E L Y P E L F D L T N H M L T L C D N I H G F K Y L I D G Y E E I N E L L Y K	168
Qy	181	L N F Y D L L R A K L D V C A N D Y C Q I P P N L K T R A N E L D V L K K L V F G Y K P L D N I K M V K M E D	240
Db	169	L N F Y D L L R A K L D V C A N D Y C Q I P P N L K T R A N E L D V L K K L V F G Y K P L D F I K D N V K M E D	228
Qy	241	Y I K N K K T I E N I N E L I E S K K T I D K N K N A T K E E K K K L Y Q A Y D L S I N K O L E E A H N L I S	300
Db	229	Y I K N K T T I A N I N E L I E G S K K T I D Q N K N A D N E E G K K L Y Q A Y D L F I N K O L Q E A H N L I S	288
Qy	301	V L E K R I D T L K K N E I K E L D K I N E K N P P A N S G W N T P N T L D K N K K I E E H E K E I K E T A K T	360
Db	289	V L E K R I D T L K K N E I K K L E D I D K I D A E K P T T G V N O I L S L R L E K E S R H E E K I K E T A K T	348
Qy	361	I K F N I D S L F D P L E Y L Y L R E K N K N I D I S A K V E ---T K E S T E P N -E Y P N G V T Y P L S Y N D I N	417
Db	349	I K F N I R L F D P L E Y L Y L R E K N K K V D V T P S Q D P T K S V Q I P K V P Y P I N G V I Y P L P L D I H	408
Qy	418	N A L ---N E L S F G D L N P F D Y T K E P -S K N I Y T D N -E R K K F I N E I K E T I K I E K K I ---E	468
Db	409	N S L A A D N K S Y G D L M N P --H Y K E I N E K I I T D N K E R I F I N N I K K Q I D L E E K N I N T K E	466
Qy	469	S D K S Y E D R S K S L I N D T K Y E K L L N E I Y D S K F P N N I D I T N F E K M M G R Y S Y E K L T H H N	528

```
QY 1 MKIIEFLCSFLPIINTOCVTHESYOELVKKLEAEDAVLTGYSLFQEKVMVNEGTS 60
Db 1 MKVIGLLSFVFAIKCKSETIEVNDIQLEKLESISVDGLEFQKSOVIIN----- 54
QY 61 AVTTSTPGSGSVASGGSGSVASGSGSVASGSGSGSGSGSGSGSGSGSGSGSGS 120
Db 55 -----ATOPTETIDPF 65
QY 121 DAKSVADIKLRVNRVLLTIKELYPQLFDLTNMLTLCNDINHGFYKLYLDGVEEINELYK 180
Db 66 TNHNF-----QQQDQVTFTEGLFTQETELVNLKALTPNRYGVKYLIESKEEFNGLMHA 122
QY 181 LNFYDLRLAKLNDYANDYCOIPNLKIRANELOVLAKLVGFKPLDNIDKNVGMKED 240
Db 123 INFYDVLRLKNDWCANNYCEIPEHLKISEBETEMLAKVILGFKPIENIQDDIEKLEI 182
QY 241 YIKNNKKTITENIELIESKTTI-----DNKNKATKEEKKKLYQAOYDLSIYNKQ 291
Db 183 YIERNKETVAALNALIAETKKIQPEGNECDNDASCDSDKYNKKKPIYOAMVNYFYKKQ 242
QY 292 LEEANHLISVLEKRDITLKKNIENIKELDKINEIKNP-----PANSQ 334
Db 243 LAEIQKVVEVLEKRVSTLKNDAIKPLWQQIEVLNAAPVWTAETQIVTGGOSSTSPGSGG 302
QY 335 N-----TPN-----TLLDN 344
Db 303 SSASCTSSGQASAGTQVQANTVASVTVTFVSGONGEASTNPQTAQVOPVPTLLEBKQ 362
QY 345 KKEIHEKEIKEIAKTIKFNIDSFTDPLELEYLUREKN-----NIDISA--KVETKESTE 399
Db 363 KKIAGLYAQIKEIAKTIKFNLEGFVDPIELEYFKKKEKESCNLSTSCCKNKASETII 422
QY 400 P--NEYPNGVYPLSYNDI-----NNALNELSPGDLINPDYTKPEKNIYTDNERKF 452
Db 423 PUTIRYNGISYPLPENDVYINKIANAAE--TTYGDLTHP--DNTPLTGDLATNEQARDL 479
QY 453 INEIKIEKIEKIESDKSKSLNDITTKIEYKLLNIEYDSKNNNIDITNFEMK 512
Db 480 IKAIKKIAKEBKLETLTNTYDNKLTFTFNQOKTFPEKAKEFYESKFNKLTSEIFEK 539
QY 513 MKKRYSYKVEKLTHTNTFASYNSKHNLKLTALKYMEDYSLRNVVYKEKLYKYNLIS 572
Db 540 KTKRDEYTKK--TELNT--CEYGTNELINKLNKQLNLYQDYSLRDKDIISNEIEYESNKK 597
QY 573 KIENIEITLVENIKDEBQLEKTKTKDENKPEKILEVSDIVKVQOKVLLMKNKIDEL 632
Db 598 ELQYINIRLAEAVQAKQNVLA-----SKDVPLSTLVELQIQKSLLTQKQIEQLN 646
QY 633 KTQLILKLVKLNHTHVPNSYKQENKQBPYYLIVLKKEIDKLKVPMPKVESLINEEKNI 692
Db 647 KTEVSLNKAQLKDLVVPKTYGNECKPEPYLIAVKVEVDRLAQPIPKIESMIAKERM 706
QY 693 K-----TEGOSDNSEPSSTEGEITGQATTKPGQOAGSA-----LEG 727
Db 707 EQGPATGESEVPSPGSAESSTDSQSTSSSSSSSTPAAABESSATLPEAPAPAEA 766
QY 728 DSVOAQAOEOKOQAPVPVPPVPEAKAQVPTPPAPNNKNTENVSKLDYLEKLYEFLNTSYI 787
Db 767 ASPSTEAEBEETIPTTQETQPSQAASSTTFAKPV-----MTKLYILEKQKLFVSYS 820
QY 788 CHKYLIVSHSTMNEKILQYKITKEES--KLSSCDPLDLFNIQNNIPVMYSMPDSLNN 845
Db 821 CHKYVLLQNSLTINKDALSXYALTSEDKIRTLKRCELDVLLAIQNNPMTYSLYESIVD 880
QY 846 SLSQLFMEIYKEMVCNLYKLD--NDKIKNLLEEA----- 879
Db 881 GLQNIYTELYKEMMYHIYKLDENPSIKSLIVKAGVIEPEVPAAPTPTVPAATBQOQOQ 940
QY 880 --KKVSTSVKLTSSSMQPLSLTPQDKPEVSANDTSHS----- 916
Db 941 ATPDVQSDAPAPSDVSQOPETPTVTSTTPEVTTSTEAASSAPGEGTPSGAGASGTGATA 1000
```

```
QY 917 -----TNLNNSLKLFENILS-----LGKNKNIYQELI 943
Db 1001 SNAATPAGTSAGSASNASNTSDVTPPAAAAAVSTSTPAPAQPAANSQSGNPDGIR 1060
QY 944 GOKSSE-----NFYEKIKDSDTFYNESTNFVYSKADDDINSUNDESKRKKLEE 992
Db 1061 SRAESEEDMPADDFELDNLYSKYQQIDG--NNTFEINFKSKKELIKALTPK--KVNLQYL 1118
QY 993 DINKLKTKTLOISFDLYNKYKLEKLERLFDKKTIVGYKQWIKKLTLLKLEQESKLSLNNP 1052
Db 1119 ETAHLKLEHSEHYDRYSTYKLEKLERLYNKHEQIQTNRQIRDLRSILKAKLRKKTQNLGV 1178
QY 1053 KHVLFNSFFNKKKEAETAETENTLENTKILLKHYKGLVYKNGESSPLKTLSESIOQT 1112
Db 1179 FYILNGYVNFNKRREAEKQYVDNALKNTDMLLKYYKARTKYFTSEAVPLKTLKASLDR 1238
QY 1113 EDNYASLENFVKLSLEKLEKLDNLNLEKKLSYLSGSLHLLIAELKEVITKNKNYNSPS 1172
Db 1239 ESNYLKIEKFRAYSRLRLKKNINLGRKISYVSGGLHHVPEEFKELIKDKDYTGKNP 1298
QY 1173 ENNTDVNNALESYKFLPEGTDTAVTVSESGSDTLEQSQPKPKPASTHYGAESNTI----- 1227
Db 1299 DNAEPTNAFEQYKELLPGKVTVST--PAVAVTTTLLAADAPATPEGAVPGAVPGAVP 1357
QY 1228 -----TTSQNVDDVDDVIIIVTFGESEEDYDOLGOVVTGEAVTPSVIDNILSK 1276
Db 1358 GAVPGAVPGSGTDTTRVAGSSVDD-----NED--DIIYQIASQSEDAPEKD--ILSE 1405
QY 1277 IENEYEVLYLPLAGVYSLKKQLENNVMTFNVNVDILNSFRNKRNFKNVLESDLIPY 1336
Db 1406 FTNESLYVYVTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQRNDPFLVLSHELDLF 1465
QY 1337 KDLTSSNVVVDPPYKFLANEKDEKFLSSVNYIKDSITDITFANDVGLVGYTILSEKYKD 1396
Db 1466 KDLSTNKVYIRNPYQLDNDKKQOIVNLKYATKGINEDIETTTDGIKFNKMWELYNQ 1525
QY 1397 LDSIKKYI-----ND--KOGENEKYPFLNNIETLYKTVNDKIDLFVHLEAKVLYNTY 1448
Db 1526 LAAVEQIATIAETANDTNKBEKKYIPILEDKGLYETVIGQAEYSBELQNRDLYKN 1585
QY 1449 EKSNEVEVIKELYNLYKTQDKLADF-----KNNNFVGTADLTSDYNHNNLLTKFLSTGMV 1504
Db 1586 EKAEEFIILTKNLEKYIQIDEKLEDFEVAENKNHASTA-----LNNLNKSGLV 1634
QY 1505 PENLAKTVLSNLLDGNLOGM--LNIS--OHQV--KQCPNSGCFRHLDERECKCLLYNK 1560
Db 1635 GEGESKKILAKML--NMDGMLLGVDPKHVCVDTRDIPKNACGFRDDNGTEBWRCLLYGK 1692
QY 1561 Q--EGDKCVENPNPTCENNNGGCDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSS 1619
Db 1693 KGEGTCTVENNPTCDINNGCDPTASQNAESTENSKKIICTCKEPTNAYVEGVFCSS 1752
QY 1620 SNFLGISFLIILMLILYS 1637
Db 1753 SSFMGLSILLIITLIVFN 1770
```

RESULT 13

A45546

major merozoite surface antigen precursor - Plasmodium chabaudi chabaudi

C:Species: Plasmodium chabaudi chabaudi

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999

C:Accession: A45546

R:Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreefs, J.; Brijs, L.; Hamers-Ca

Mol. Biochem. Parasitol. 43, 231-244, 1990

A:Title: Molecular cloning and sequence analysis of the gene encoding the major meroz

A:Reference number: A45546; MUID:91218805; PMID:2090945

A:Accession: A45546

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1785

A:Cross-references: GB:M34947; NID:g160597; PID:g160598

C:Superfamily: major merozoite surface antigen

C;Keywords: surface antigen

Query Match	29.88;	Score 2509;	DB 2;	Length 1785;
Best Local Similarity	31.64;	Pred. No. 1.5e-70;		
Matches	598;	Conservative 324;	Mismatches 605;	Indels 366; Gaps 35;
QY	1	MKIIFLCSLFFIIITQCVTHESYQELVKKLEALEDAVLTSYLFQEKMLVNEGTS	60	
DB	1	KAIGLFSFVFYAIYCKSETIGYVNDLVHKLKLELSEGLELFQKQVIVNAQSPET	60	
QY	61	AVTTSTPGSGSVASGGSGSVASGGSGSVASGGSGVSGSGNSRRNTSPDSSDS	120	
DB	61	PVDPF-----TNP-----	68	
QY	121	DAKSYADLKHRVRYLLTIKELKYPQLFDLTNNHMLTLCDNTHGPKYLDGTEENELLYK	180	
DB	69	-----EFAQQLFQILFKELSGTEQTEGLVNLTKTLGPNKYGLKYLSEKEFNELMHA	122	
QY	181	LNIFYDLRAKLDVANDYCOIPFNLIKIRANELDVLLKLVFGYKRPDLNKTGNVGMED	240	
DB	123	INFIYDVLROKLDWCANNYCEIPEHLKINVEIEMLKVVLYGKYPLENLODDLKLEE	182	
QY	241	YIKKNKTIENINELIESKTI-----DKNKNATKEEEKKLLQAOYDLSIYNKK	291	
DB	183	YIARNKATAETPLNLTIEETKKITPEEBTDCNDTNCNTKYGKKKAIYQAMYNVIFYKKQ	242	
QY	292	LEEAHNLISVLEKRIDTLKKNENIKELLDKINEIKNPPANS-----	334	
DB	243	LAELKKVIEVLEKRVATLKKNKAEIKPLQQLTEALRGPPAVTEGQIATGSGSEETKQNSTE	302	
QY	335	--NTPNTLLD-----	342	
DB	303	SSNTKTTTDDKAVTTQATKATGTETNTGTETGTETNTATGTTATGTTATGTPTVT	362	
QY	343	-----KUKKTEEHEKEKEIAKTKFNIDSLFTDPLELEYLRE-KNKNI	386	
DB	363	EPVOVPAVQVLTEBEKAKKTAELAQKIEAKTKFNLDGIFVDPVELEYLYKKKKNESE	422	
QY	387	DISA---KVETKESTEP--NEYPNGVYPLS---YNDINNALNELNSFGDILNPFVDYTK	437	
DB	423	HSTSSCHKNTPEIVPLNVRYPNGISYPLTEEVVYSKIAHNAE--TYIGDLTN-VDNTA	480	
QY	438	EPSKNITDNRKFFINEIKKIEKKIESDKKSYEDRSKSLNDITKEYEKLKLEIYD	497	
DB	481	ITEDLTTNEQAKNLIKAIKKIEAEQKVELDQDDYTKLAAPNGQRTPEEAAKKFE	540	
QY	498	SKFNNDLDTNFKXMGKRRSYKYVEKLTHHTNTEFASYSKKNLEKLTAKALYMEDYSURN	557	
DB	541	SKFKNLTTDIFDFKRYKRYEMNKKAA--LVGCEYGNTOQLINKLNQLNLYLDYGLRK	598	
QY	558	IVVEKLUYKNLTSKIENEITLVENIKKDEOLFEEKITIKDENKKPDEKILEVSDIVKV	617	
DB	599	EIVNTEIFYSNKSEQLQYNINRLANVAQKNTLVA-----SKHPLSTLVEL	647	
QY	618	QVQVLLMNKIDELKTKQLLLKNVLKHNHVPNSYKOENKOEYPYLLVLKAKETDKLVF	677	
DB	648	QIQSLTKLTEQNKTEFSLNKAHLKDKIYVPTYGKEGRPEPYLLIAIKKIDRLAKF	707	
QY	678	MPKVESLINEKKNIKTEG-QSDNSEPSTEGEITGQATTKPGQQA-----	721	
DB	708	IPKIDDMIEKEQKQKEHVATGESQASSAGTSGSTETTTSQAPAVPAAPAEAKAKE	767	
QY	722	-----GSALEGDSV---QAAQOQKQAO-----PVPVPEAKAQVPT	757	
DB	768	GTESTEETPAASKPAEGAASATGATPTTEQEAAPTQEAQAPAEPTPAEPAPTTPAAPAT	827	
QY	758	PPAPVNNKTEVSKLDYLEKLEFNTSYICHKYLIVSHSTMENKILKOYKITTEES--	815	
DB	828	PAAAPAAKPVMTKLYLEKLLKELAFASYACKKVVLLQNSTINKDALSKYALTPEEDKIR	887	
QY	816	KLSSCDPLDLFNTONNIPVMYSFDSLNNLSLSOLFMEIYEKEMVCNLYKLKD-NDKTN	874	
DB	888	TLKRCSEDLVLAOTNNPMTYSLYENVVDGLONITVELYEEKMMHYHLYNLDKNPVAK	947	

Qy	875	LL-----	-----BEAKVSTSVKT- 888
Db	948	LLKAGVIDPEVPAPETAPETAPETAPETAQEAPOQESAQAPEATETTP 1007	
Qy	889	LSSSMQPLSLTPQDKP-----	-----EVSANDTSH 915
Db	1008	AESASTEPTKAPTATPTSETVTQEGTTPAAPKAQEGASSAPAPAKPAPATQVTGQ 1067	
Qy	916	STNLNSLKLFENILSLGNKNKYBELIGOKSSENFYKILKDSDFYNESFTNFVASKA 975	
Db	1068	STNVEGSTQV-----	-----RAESEDENFVDDFEVDNFYKSYLQOVDG-NTQFIDFIKSK 1118
Qy	976	DDINSLNDESKRKLBEDINKLTKTLQLSFDLYNKKYLKLERLFDKKTVGKYKMQIKKL 1035	
Db	1119	ELINALTPE-KVNQLYLDIAHLKELSEHYNNRYKYKLERLYQKHOIEAANKVKKEI 1177	
Qy	1036	TLIKEQLESKLSLNNPKHVLFQNSVFFNNKKEAELIAETENTLENTKLLKHXYGLVKYY 1095	
Db	1178	SVLSRLLKRBKKYINGTFYVLSGFANFFNKRREABKQVVDNAIKNTDMLLKYYKARSYF 1237	
Qy	1096	NGESSPLKLTSEESIOTEDNYASLENFKVLSLCKGLKLDNLNLEKKLSYLSSSLGHLHIA 1155	
Db	1238	TSRAVPLKLTUKTYSIDREANYLKEIFRAYSLERLKKNLINLKERITYYVSGGLHWFPE 1297	
Qy	1156	ELKEVINKNKYTGNPSENNDVNNALESYKPLPEGTD-----	-----VATVVSSEGSTLTLEQS 1210
Db	1298	EKELLNKGQYTKTNPENAPEVIRAFEOYKELLPKGATTPAPVVPVAPAPATAAPAA 1357	
Qy	1211	QPKKP-----	-----ASTHVGAESNTTTTSONVDDEVDVVIPIFGESEEDYDGLQ 1258
Db	1358	DAPVAAAAAASGSGSAATTEGEAATTVVASDND-----	-----DDDDMDQ 1402
Qy	1259	VVTGEAVTPSIDNILSKTENEVEVLYLAKPLAGVYRSCLKOLENNVMTFNVNVDILNSR 1318	
Db	1403	IANAQS-TDEVDKIIDAFKASENYIYTKSLGNTYKSKFKHMLKEFMIKEDIWGLNYK 1461	
Qy	1319	FNKRENFKNVLESDLPIYKDLTSSNVVVDKPYKFLNKKRKFKESSYNYIKDSIDTDINF 1378	
Db	1462	LEKRNFDLVSYELAFDXDINTNKFVKNPYQLLNDKKDKQMINLKYAIKGVTEDIET 1521	
Qy	1379	ANDVLGYKILSEKYSDDLSDKKYI-----	-----NDKOGENEKYLFPFLNNIETLYKTVNDKID 1433
Db	1522	ATDGIFFNKMIELYKPOLNAVNEQIAAIGTEPTDAEKKKYAPFEDULGLYETILNGAE 1581	
Qy	1434	LFVTHLEAKVLNVTYKSNVEYKIKELNLYKTIODKLADF-----	-----KKNNFVYGIADLSTDY 1489
Db	1582	EFSELQHKLENTKIBKAGFDILMANLEYIRIDKELEDFVESAEKKNHIASIA-----	----- 1635
Qy	1490	NHNNLLTKFLSTGMVFENIAKTVLSNLLDGNLQGLMINT-SHQCVKYKQCPONSGCFRHL 1548	
Db	1636	-----LNNLKSGLVTEGESKKILAKMLNMDAMDLLIGSNHVCISTSTPDNAGCFRYDD 1690	
Qy	1549	ERRECKLLNYKQ--BGDKCVENPNTCNENNGGCDADAakteedsGSN--GKKITCECT 1604	
Db	1691	GTEWRCLLGGFKDDGNRCVADDAVPCVNMMNGGCDKNADCREVENTDRDPSKIVCTCK 1750	
Qy	1605	KPOSYPFLDGIFFCSSNFFLGISFLLLILMILYS 1637	
Db	1751	EPNPNNYAGVFCSSGFMGLSILLIITLIVEN 1783	
RESULT	14		
		547282	

RESULT 14
S47282

merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71)
 C:Species: Plasmodium falciparum
 A:Variety: strain RO-71
 C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
 C:Accession: S47282
 R:Tolle, R.; Bujard, H.; Cooper, J.A.
 submitted to the EMBL Data Library, July 1994
 A:Description: Plasmodium falciparum: recombination within the C-terminal region of m
 A:Reference number: S47282

A:Accession: S47282
A:Molecule type: DNA
A:Residues: 1-651 <TO>
A:Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A:Experimental source: strain RO-71
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 25.5%; Score 2146; DB 2; Length 651;
Best Local Similarity 64.5%; Pred. No. 8.3e-60;
Matches 422; Conservative 82; Mismatches 114; Indels 36; Gaps 5;

QY 1019 FDKKTVGKYMQIKKLTLLKQESKLSLNPNKHVQLNFSVFNKKKEAIEAETNTL 1078
Db 1 FNKKELGQDKMQIKKLTLLKQESKLSLNPNVQLNFSVFNKKKEAIEAETNTL 60

QY 1079 ENTKILLKHYKLVKYNGESSPLKTLSEESTQIEDNTVASLENFKVLSKLEKLDNLNL 1138
Db 61 ENTKILLKHYKLVKYNGESSPLKTLSESVTQEDNTANLEKFKALSKIDKGLKNDNLHL 120

QY 1139 EKKKLSYSSGLHLIAELKEVIKNKNTGNSPSENNTDNNALSEYKFKFLPEGTDVATV 1198
Db 121 GKKSFLSSGLHLITELKEVIKNKNTGNSPSENKKVNEALASYENFLPE-AKVTV 179

QY 1199 VSE-----SGSDTLEOSQPKKPASTHVGAESENTITTSQNVDDDEVVLIIVP 1244
Db 180 VTPPQDVTPSPLSVRVSGSGSTKEETQIPTSGSLLTLEQOVVOLQNYDEDDSLVVL 239

QY 1245 IFGESEEDYDGLGOVVTGEAVTPSIDNLSKIENEYEVLYLKLPLAGVYRSLLKQLENNV 1304
Db 240 IFGESEDDNDEYDQVVTGEAISVT-MDNILSGFENEYDVIYLDKPLAGVYRSLLKQIEKNI 298

QY 1305 MTFNPNVADILNSRFNKNRENKFNLESIDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSS 1364
Db 299 FTFENLNDILNSRLKKRYFLDLESQDMQPKHSSNEYIIEDSKLLNSQKNTLKS 358

QY 1365 YNVIKDSIDTDINFANDVLGYKILSEYKSDLSIKYI-----N 1405
Db 359 KYIKESVENDIKFAQEGISYEVYKYLAKYKDDLESIKKVIKEKEFPSPPTTPSPART 418

QY 1406 DKOGENEYLPPLNIETLYKTVDNKIDLFTVHLAKVLNVTYKESNVKIKELNYLKT 1465
Db 419 DQKKESFELPPLNIETLYNNLVKIDYLLNLAKINDCNVKEDEAHVKITKLSDLKA 478

QY 1466 IODKLADFKPNNFVGIADLSDYNNHLLTKFLSTGMVFENLAKTVLSNLDGNLQGLM 1525
Db 479 IDDKIDLKFNHNDPEAIKKLINDTKDKMLGKLLSTGLV-QNFPNTIISKLEKGFQDML 537

QY 1526 NISQHCYKVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADA 1585
Db 538 NISQHCYKVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADA 597

QY 1586 KCTEEDSGSGKKITCETKPDSPYPLFDGIFCSSSNFLGISFLLLMLILYSFI 1639
Db 598 KCTEEDSGSGKKITCETKPDSPYPLFDGIFCSSSNFLGISFLLLMLILYSFI 651

RESULT 15
A45545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments)
C:Species: Plasmodium falciparum
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C:Accession: A45545
R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1
A:Reference number: A45545; MUID:92131048; PMID:1775158
A:Accession: A45545
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-400 <BLA>
A>Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBIIP:77621)
C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 24.9%; Score 2101; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 1.2e-58;
Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1240 VIVPIFGESEEDYDGLGOVVTGEAVTPSIDNLSKIENEYEVLYLKLPLAGVYRSLLKQ 1299
Db 1 VIVPIFGESEEDYDGLGOVVTGEAVTPSIDNLSKIENEYEVLYLKLPLAGVYRSLLKQ 60

QY 1300 LENNVMTNVNVKDTILNSRFNKNRENPKVLESIDLTPYKDLTSSNVVVKDPYKFLNKEKRD 1359
Db 61 LENNVMTNVNVKDTILNSRFNKNRENPKVLESIDLTPYKDLTSSNVVVKDPYKFLNKEKRD 120

QY 1360 KFLSSYNYIKDSIDTDINFANDVLGYKILSEYKSDLSIKKYINDKOGENEKYLPLFN 1419
Db 121 KFLSSYNYIKDSIDTDINFANDVLGYKILSEYKSDLSIKKYINDKOGENEKYLPLFN 180

QY 1420 NIETLYKTVDNKIDLFTVHLAKVLNVTYKESNVKIKELNYLKTIODKLADFKNNNF 1479
Db 181 NIETLYKTVDNKIDLFTVHLAKVLNVTYKESNVKIKELNYLKTIODKLADFKNNNF 240

QY 1480 VGIADLSDYNNHLLTKFLSTGMVFENLAKTVLSNLDGNLQGLMNTISQHCYKVKQCPQ 1539
Db 241 VGIADLSDYNNHLLTKFLSTGMVFENLAKTVLSNLDGNLQGLMNTISQHCYKVKQCPQ 300

QY 1540 NSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADAKCTEEDSGSGKKI 1599
Db 301 NSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADAKCTEEDSGSGKKI 360

QY 1600 TCECTKPDSPYPLFDGIFCSSSNFLGISFLLLMLILYSFI 1639
Db 361 TCECTKPDSPYPLFDGIFCSSSNFLGISFLLLMLILYSFI 400

RESULT 16
A25814
glycoprotein 185 - malaria parasite (Plasmodium falciparum) (isolate Honduras I)
C:Species: Plasmodium falciparum
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jun-2000
C:Accession: A25814
R:Howard, R.F.; Ardeshir, F.; Reese, R.T.
Gene 46, 197-205, 1986
A:Title: Conservation and antigenicity of N-terminal sequences of GP185 from different
A:Reference number: A25814; MUID:87106855; PMID:3542719
A:Accession: A25814
A:Molecule type: mRNA
A:Residues: 1-196, 'T', 198-233 <HOW>
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein

Query Match 14.1%; Score 1187; DB 2; Length 233;
Best Local Similarity 95.1%; Pred. No. 1.2e-30;
Matches 232; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 1 MKIIFFLCFLFFIINTQCVTHTESYQELVKKLEALEDAVLTGYSLFQEKMYLNEGTSQT 60
Db 1 MKIIFFLCFLFFIINTQCVTHTESYQELVKKLEALEDAVLTGYSLFQEKMYLNEGTSQT 60

QY 61 AVTTSTPGSKSVASGGSGSVASGSGSVASGSGSVASGSGSGNSRRTPNPSNDS 120
Db 61 AVTTSTPGSKSVASGGSGSVASGSGSVASGSGSGNSRRTPNPSNDS 108

QY 121 DAKSYADLKHVRNVLTIKELKYPQLFDLTNHLMTLCDNIHGFKYLDGYEETINELLYK 180
Db 109 DAKSYADLKHVRNVLTIKELKYPQLFDLTNHLMTLCDNIHGFKYLDGYEETINELLYK 168

QY 181 LNFYFDLLRAKLDVNCANDYCOIPFNKIRANELDVLKLVFGYKRPDNTKDNVGMED 240
Db 169 LNFYFDLLRAKLDVNCANDYCOIPFNKIRANELDVLKLVFGYKRPDNTKDNVGMED 228

QY 241 YIKK 244
|||||

QY 235 VGRMEDYIKRNKKTENINELI-EESKKTIDKNKNAK-----Z/I

QY 235 VGRMEDYIRKKNRKTIENTINELI-EESKRTIDKKNKNATK-----

Db 1829 INTLOTKYDLNQIN-----KRVSMVADATNDNNLNLIKKEKEATKTINNLT 1875
QY 1573 TCNENNGGQAD 1584
Db 1876 LFTIDSNNKIDAD 1887

RESULT 25
T28677
rhoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A:Reference number: 220508; MUID:95021522; PMID:7935623
A:Accession: T28677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KEE>
A:Cross-references: EMBL:127838; NID:9457145; PID:9457146; PIDN:AAA21304.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a plasmodium yoelii rhoptry protein. Multiple c
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match 6.7%; Score 568.5; DB 2; Length 2269;
Best Local Similarity 21.4%; Pred. No. 2.1e-10;
Matches 384; Conservative 248; Mismatches 534; Indels 625; Gaps 81;

QY 166 YLDIGYEINELLYK-LNFYFDLLRAKLDVNCANDYCOIPFNLIKIRANELDVL--KKLVF 222
Db 155 YKVTGYENKNTIYNTIKSYFDQIYEGDIDTFYNELSSI-----VKREDPIDIEDTKLE 209
QY 223 GYRKLPLDNTKDNVGRME-DYIKKNNKTYIENINELIEESKTTIDKNNKAYKEKKLYQA 281
Db 210 NLRSDKINDYDKTQMEIETVKSHLNINETNNK-----PNTILEIKKIYIDEISKEL--- 262
QY 282 QYDLSIYNKOLEAHNLISVLEKRI-DTLKKENIKELDKKINEIKNPPANSNTPTNL 340
Db 263 -----NKMLEDFKKEKELSKNSIDYDKKREQLSEYKSKMLEIRNHVNSQT-NVDNTK 314
QY 341 LDKNK-----EKEIKEIARTIKFNIDSLSFTDPLELEYLYREKNNKIDIS 389
Db 315 EEEAKQNYDKSNEHMTIPTNDEISKIISVKTMDKI-----LSKYNTYIDFN 364
QY 390 AKVETKESTPEPNVGVYPLSYNDINNALNELSGDLINPFDYTKPESKNIYTDNER 449
Db 365 KKY--KET-----VNSHSQFTELTDKIKAEVSDKELKCEQSFNDN 404
QY 450 KKFINEIKKIKTEKKIESDKK--SYEDRSKSLND-IRKEYEK-----LLNE-IYDSK 499
Db 405 KSLINETKSIEKEYQINFLKVDYETIKVCKSTKESITKFSKQITLKDMLNQNTKVK 464
QY 500 FNNIDLTNFEK-----MMGKRRYSKYVEKLTHTNTFASYENSKHNLEKLTAKALYMEDYS- 554
Db 465 ETNSIDKSIEKEQIILTKQ-----TKLENKFTETSLANNHEANNEL---IKYFSDLKA 516
QY 555 -----LRNVVKEKLY-----YKNL-----ISKIENIETLVENKTKDEE----- 590
Db 517 NLGINBENMLNYOFTKEKTFNDIKEKNIHINEISKIEIKIHASINYINSEETEREIGIN 576
QY 591 -----OLPEK-----KITKDKENKPKDEKILEVS--- 612
Db 577 IESLNTKVFKEKVENVTNLNKIEKLUKHYDFSGEGNIKIYTDKIKKINDDIMAVSQOI 636

QY 613 -----DIVKQVQOKVLLM-----NKIDEL-----KKTQILKLNVELKHN 646
Db 637 DQHINGLDDIQKKSSEYVSEMKEQINKLEKVSNTESNDNVEGIKKKQIIVTKDKKN 696
QY 647 IHVPNSYKQENKOEPPYLLIVLKKKEIDKLVFMPKVESL-----INEEKN 691
Db 697 I-----YEEINK-----LLSEISKIEKDNSTLEKVDNLSYQGNLGNLFLEQIDEKK- 745
QY 692 IKTEGSDNSEPSTEGETGQATTKPGQAGSALDEGDSVQAQEQKQAPPPVPVPEA 751
Db 746 -----KAENTIKSWEAYID-----DLDNKKKSQE----- 770
QY 752 KAQVPTPPAVNNKNTENVSCLDYLEKLYEFLNTSY-----ICHKYILVSHSTMMEKILKQY 807
Db 771 -----IETEMDIKMD-INKEMEALKISHDDDKKCHKSKNHNKENISDIYDKSS 817
QY 808 KITKE--ESKLSLSCDPLDLLFNQIIONIPVWYSMFDSLNLSLQLPWEIYEKEMWCNLYK 865
Db 818 KIIDFSRESIDN-----DIKNKLQKNV-----SESONHNSDINOCLNE-----VANIYN 862
QY 866 LKDNKIKNLEEAKKVSTSVKTLSSSSMQPLSLTPQDKPEVSANDDTSHSTNLNLSKL 925
Db 863 ILKLNKIKKIIDKVKETSEI-----EKNKKNINDE-----LNNSEKV 900
QY 926 FENI---LSLG-----KNKIYQELIGQ-----KSENFYEK 954
Db 901 IKKIEGDLSLKECRKINSKINSTLDDKDIECINKINVLKKNILNEETNITNHFKAEE-YNK 959
QY 955 I-----LKSDTFY-----NESFTNFVSKADDIINSLNDESKRKKLEEDINKLKT 1000
Db 960 IVLSNFNNIEMADNKSQYILEIKKNGTNDHDYNIKELSHKDKSNQYKTEADQN---KKA 1017
QY 1001 LQLSFDLYNRYK-----LKLERLFDKKTKVGYKMQIKK-----LTLKE 1040
Db 1018 IQKNKELFEQYKEEVTVLLNKYAVELKNFKDKTKNDKSK--OIIRKIDAHNYCTLESG 1074
QY 1041 QLESKLNSLNPK-HVLQNFVSFFNKKKAEIAETENT---LENTKILLKHYKGLVKYYN 1096
Db 1075 KSEKMKNEIKNEKIHI-----EDEVANDKSNKAITSIKVSVPFFTKIKIN 1122
QY 1097 -----GESPL-KTLSEESIOTEDNYASLENFVKLSLE-----GKLKDNLNLE 1139
Db 1123 EIRTKSDDCLEKFNDEKQLSNLSIDTQETPLT-ENGKQLKTLEELLESUKKOKKNIEDQ 1181
QY 1140 KKLSYLSLSSGLHH-----LJAEKLEVIK-NKN----- 1165
Db 1182 KRELDEVNSKIKNIENVTNQHKKNYEIGIVEKINEIAKTNNKQIESTKELIKPTIQIHS 1241
QY 1166 -----YTGNSPEN-----NTDVNNALESYKFKLPEGTDTVATVVSSEGSOTLQSOPKPK 1215
Db 1242 SFNANDLEGIDSDELKGYNTMGNIYEETIKSYNLITNYLETVSKE-SITYNQIQNKR- 1299
QY 1216 ASHVCAESNTIITSQNV-----DDEYDDVI-----IVPFGSESE 1251
Db 1300 -----IDTQKELLKNIEENVNNAKSYLDYIKENEFDRIVTHFKKLNTVNDNFKNEYSKVNE 1355
QY 1252 DYDDLQGVVTGEAVTPSDVIDNLSKIENIEVEVLYLKLPGAVYRSLKKQLN-----NV----- 1304
Db 1356 GFDNISNSI--NTVKNTSDENSLNLLNQTKEMYANIVNTNYYSYKYEAEINERNIPKLA 1413
QY 1305 MTFNVNVKD-----ILNSRFNKNRKNVLESDDLIPYK 1337
Db 1414 NTLNIKIKNSSGIDLSKDIKIALSYLDSKTEDTLIFIPSPQKKTETTYTISDSYSILLD 1473
QY 1338 DLTSSNVVVKDPKFLN-----KEKROKFLS----- 1363
Db 1474 ILKKSQELQKKEOQTLKIFENRRLYEKVOATNELRGTTSLDKYKKEKILSEVKKLLHKS 1533
QY 1364 -----SYNIKDSIDTIDINFANDVLGYKILSEKY 1393
Db 1534 NELNKLSCNFQNYDTILESSKYDQVKEKSNNYKQEKELGIDFNVD-----MEEKF 1585
QY 1394 KSOLDSEIKKIYIN--DKOGEN-----EKLPLPLNNIETLYKTIVNOKI-----DLF--VIH 1438

Db 1586 NNDIKVIELENNYSSSENNILQSKQKLELTNKAETKIDDKIDKIERNDLIDKLE 1645
Qy 1439 LEAKVLNITYEKSNEVEKIKELNYLKTIDQKLADFKKNNFVG-----IADLSDY 1489
Db 1646 TRKNCMLFTHTLAETLIKIDTYSKFIESATKFSKFLKYPGTFNSLNDIDIAFLQKY 1705
Qy 1490 -----NHNLLTKFLSTGMVFENLAK--TVLSNLLDGN 1520
Db 1706 DLHQINKYVTSKLSDATNDNNLIEKEATQAINKLTKLTIDSNNIDAN 1756

RESULT 26
C71622
hypoetical protein PFB0145c - malaria parasite (Plasmodium falciparum)
C:Species: plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: C71622
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravid, L.; Koonin, E.V.;
Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: C71622
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1979 <GAR>
A:Cross-references: GB:AE001375; GB:AE001362; NID:g3845105; PIDN:AAC71819.1; PID:g384510
A:Experimental source: clone 307
C:Genetics:
A:Gene: PFB0145c

Query Match 6.6%; Score 559.5; DB 2; Length 1979;
Best Local Similarity 22.2%; Pred. No. 3.4e-10;
Matches 370; Conservative 314; Mismatches 609; Indels 373; Gaps 82;

Qy 110 RTNPSDNSSDAKSYADLKHVRNYLTITIKELYPQLFDTNHMLTCDNIHGFYIUD 169
Db 126 KDNELENQDKTLKSISSLNKIVYESKIEBLE-KELKEVKDKNI-----N 173
Qy 170 GVEETNELLYKLFY--FDLLRAKLDVNCANDYCIQIPFNKIRANELDVLKLVGYRK 226
Db 174 DYE--NKLKEKDFVKQIDMLNEKEN-----LLQKEKELDINKR-----EK 212
Qy 227 PLDNKIDNVGKME-----DYIKKKKTTIENINELIEESKKTIDKNATKEEKKKL 278
Db 213 KINEKKNIIKKEETFNIEKYLEKNER-ETISIEIIDIKKHLEKLEIEKKE-- 268
Qy 279 YQAQVDSIYNKQLEAHNLISVLEKRDITLKKNIKELDKINEIKN-----PPPANS 333
Db 269 -----DLENLAKKLSKENVLKEKGVK--EKNETINSLANDNIIEKKEKYLLEYEE 321
Qy 334 GNTPTLDKNNKIEEKE--IKEIAKTKFNIDTSLDTPLELYYLRKKNKIDT-S 389
Db 332 KKKQIDLNKQKE 381
Qy 390 AKVETKESTEPNXPNGVTPSYNDINNALNEL-NSFGDLINPDYTKPSKNITYDNE 448
Db 382 MDIEKREH-----NFLHMDQLKLNKSNFVKNNNQKLVYKCEIKNLKTELE 427
Qy 449 RK-----KFNIEKEKTKIEKKTKESDKKSYEDRSKSLNDITKYEKLL 492
Db 428 KKEKELKDIENYSKEEINKLNQNEK----EKQILAFNKNHKEIHGLKEELKESVR-I 482
Qy 493 NEIYDSKFNNDITLNEE-KWNGKRYSVKVEKLTHTHTFASVENSKHNLKLTALKAYE 551
Db 483 TKIETQELQEMVDIIOKELDQLEKYNQIESI-----STELSKKEKYNQYKNTIE 535
Qy 552 DYSLRNIVVEKELKYKNL-----ISKIENIEITLVENIK-----KDEQLPEK 595
Db 536 EINNLEKLEETNKETNLONNYTNEINLNDIHLNAGNIKMTQISTLKNVDHLLNE 595
Qy 596 KITDENKP-----DEKILEVS-DIVKVQVQKVLMMNKIDEL-KKTOLILKNVELKHN--IH 648

Db 596 QIDKLNNKGLTNSKISLBNVQIMDLKEKDFLNQIDVLSNQIDLLTRKMEKENKMLE 655
Qy 649 VPNSYKQE-----NKQEPYLIJVLKKEIDKLKVPMPKVESLINEKKNITKTEGSDNSE 702
Db 656 QENKYKQEMELLRGNIKSSENILNNDNEVCDLK-----RKLKSEKEMKMMKEEHDKLAE 711
Qy 703 PSTGEITGQATTTPGQAGSALE---GDSVQAQAEQKQAPVPVPVPEAKAQPVP 759
Db 712 LKDCDVRIRMEKNEKDKINMLKEEYEDKINTLKEQNE----- 751
Qy 760 APVNN-KTENVSKLDYLEKLYEF-LNT--SYTCHYKILVSH-----STMNEKILKQYKI 809
Db 752 -KINTLKEQNEKDKINTLKEEYEHKINTMKEEYEHKINTLNEQNEHKINTLNEQNEHKINT 810
Qy 810 TKEE-ESKLSKCDPLDLFNQNNIPVYSFDSLNLSLSOLF--EYKEMVNLKYL 866
Db 811 MKEEYEDKAMTIN-----EONEDKNLSKKEEYENKINQINSNEIKIDV-NEV-I 860
Qy 867 KDNKIKNLEEAKE--VSTSVKTLSSSQPLSLTPQDKPEVSANDDTSHSTNLNLSL 923
Db 861 BEVDKLVTLDEKKKQDFDKINEYAHIKAEKEQILLTEME--ELACQDNKYSDLYEKYI 918
Qy 924 KLFENILS-----LGNKNITVQELIGQKSSSENFYKILKDSOT 961
Db 919 KLIKSIEMIITECDDDIENEDIIIRRIEYINNNKGLKEVEEKEHKHSSFNILSKREK 978
Qy 962 FYNESFTNFVSKADDDINSLANDE--SKRKKLEEDINKLKTQLSFDLYNNYKILKLERL 1018
Db 979 F----FKNSIEDKSHELKKKHEKDLLSKDKEIEE--RNKKIKELNNDI--KKLQDELL 1028
Qy 1019 FDKKTGVKYMOKIKKLTLL-----KEQESKLSNLPKPHVLQNFVSFFNKKKAEIAE 1073
Db 1029 VYKGSNAQVDDHKKSWILLKDKSKEIKDKENQINVEKNEEKDL-----KKKDEI-- 1081
Qy 1074 TENTLENTKILLKHVYGLVKY-----YNGESSPLKTLSEESTQTEDENYASLENFKVLKLE 1129
Db 1082 -----RILNEE--LVKYITILYNLKDPLQONQDLSKIDINSLTINMGCMVDKIE 1130
Qy 1130 GKLD--NLANKKLSVLSGLHLLAEKLEKIVKNNYTGNSPENN-----TD 1177
Db 1131 EHILDYDEINKSRNSLQKNEICSLTEVWELNKKN--ELIEENKLNLDVQKKK 1187
Qy 1178 VNNALESYKFLPECTDVATVVSESGDTLEQSQPKKPASTHVGAESNTFTTSQNVDDV 1237
Db 1188 LKQVQKQKETEKLKQKNTKCNQIDELNEVEKLNENIELITYSNDL---NNKFDMK 1244
Qy 1238 DDVIIVPIGSEEDYDDLQGVVTGEATVPSVIDNLSKIEN-EYEVLY---LKPLAGV 1292
Db 1245 ENNLMMKL-----DENEDN-----IKMKSKIDDMKEKEIKYREDEKRNUNE 1286
Qy 1293 YRSKKQLENVMTFN-VNVK--DILNSRFNKNRFKNV-LESDLIPYK-DLTSSNYVVK 1347
Db 1287 INNLAKKEDMCIKYNEMNIKYGDICVKEYSMTYKTSKYEQIKVYDEKCSQY--- 1343
Qy 1348 DPYKPLNKEKDKF-----LSSYNYKDSIDTDINFANDVLYGYKILSEKYSOLD 1398
Db 1344 DEIRFOYDEKCFQYDEINKKYGALLNITNKWDSKVDNRNN-----EIIISVDNK--VE 1396
Qy 1399 SIKKYINDKQGENEYLPF---LNNIETLYKT-VNDK--IDLFIHLEAKVLNITYEKN 1452
Db 1397 GIANYLQOIFELNEEIIIRLKEGINKISLYSNELNEKSYDINMKHIOBQLL--FLEKTN 1454
Qy 1453 VEVIKELNLYLKTIDQKLADFKKNN-----FVGTDLSTDYNNHNLTKFLSTGV 1504
Db 1455 KENEKIINLTSQYSDA---YKKKSDSEKLCGAQFDDVNTYGNISNNIRNTEYKYEEM 1511
Qy 1505 FENLAKTVLSNLLDGNLQGLMNLISQHCYKQKOPNSGCFRHLDEEKC-----LLN 1558
Db 1512 FD-----TNEEKN--GM-HLSKYIHLEE--NKFRCKMKIYENENIKSSNKIIGLYN 1559
Qy 1559 YKQ-----EGDKC-----VENPNTCNENGGCDA-DAKCT 1588

Db 1560 YSRYYGLREDLCKEEI VPSKIGNISKNENNKKNNNTCDGYDEKVT 1605

RESULT 27

B25120

C;Species: Plasmodium falciparum
C;Date: 20-Aug-1987 #sequence_revision 20-Aug-1987.#text_change 09-Jun-2000
C;Accession: B25120; A05263
R;Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Stunnenberg, H.; EMBL J. 4. 3823-3829, 1985
A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium falci
A;Reference number: A91030; MFID:86136024; PMID:3004972
A;Accession: B25120
A;Molecule type: DNA
A;Residues: 1-115 <MAC>
C;Superfamily: major merozoite surface antigen
C;Keywords: merozoite; surface antigen; tandem repeat
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-115/Product: major merozoite surface antigen #status predicted <MAT>
F;69-101,108-113/Region: 3-residue repeats

Query Match 6.38; Score 529; DB 2; Length 115;

Best Local Similarity 91.7%; Pred. No. 1.2e-10;

Mismatches	Conservative	Mismatches	Indels	Gaps
1111	0	4	4	6

QY 1 MKIIFFLCSELFFIINTQCVTHESYQELVKKLEALEDAVLTCYSLFQKEKMLNEGTS 60

DB I MKIIFFLC SFLFFIINTQC VTHESYQELVKKLEALED AVL TGYSLFQKEK MVLNEGTSGT 60

QY

A vertical ruler with markings from 0 to 10 cm. The markings are in millimeters, with major ticks every centimeter and minor ticks every millimeter. The ruler is oriented vertically with the 0 mark at the top.

Db 61 AVTTSTPGSSGVT---SGGSVASVASGG---SGGSVASGGGNSRRTPNSDSSDS 11

Ov 121 D 121

7

Db 115 D 115

RESULT 28

T18501

hypothetical protein C0/60C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum

C/Date:	15-Oct-1999	#sequence_revison	15-Oct-1999	#text_change	09-Jun-20
C/Date:	15-Oct-1999	#sequence_revison	15-Oct-1999	#text_change	09-Jun-20

C;Accession: T18501

R; Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library August 1997

Submitted to the ERIC Data Library, August 1997
A; Reference number: Z18935

A;Accession: T18501

A; Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: DNA

n, molecule type: DNA
A; Residues: 1-3394 <LAW>

A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB1114

C:Genetics:
A:Map position: 3

A; Map position: 3
A; Note: C0760c

Query Match	6.28;	Score 521.5;	DB 2;	Length 3394;
-------------	-------	--------------	-------	--------------

BEST LOCAL SIMILARITY 20.8%; Pred. NO. 9.3e-09;
Matches 418: Conservative 299: Mismatches 604:
Indels 685: Gaps

[illegible]

QY 108 SRRTPNSDSSDAKSYADLKHRVRNYLLTIKELKYPQLFDLTN-----HMLTL----- 15

613 SHNNKSDYDEGDAEN-----NNVYTKTKKNKYKKNLNLNKAKMKITTLTRANEM 66

RECEIVED

Qy 158 ---CDNIHGF-----KYLIDGYEEI-----NELLY-----KLNFYFDLLRAKLN DV 19

6'5 EFKYC\$NTKNTI.TRODMKKERKPRISOVHIIHNEKIYIEKLI NEKINVTKD-TEKKIDEL. 72

—

Qy 196 --CANDYCQIPFNKIRANELDVLLKLVFGY-----RKPLDNIKDNVVK 23

724 HGVNKNKEDTYIIQVEKOTI.IKVTSVVVDYTKMESENHIEKMTTWNKMI.NNV--HMSS 78

07 SCHEI AMELIANNY TANNEN¹ TITANESCELYV TQY A SCCT AYTETZUWLA BERTTTCGNNUNNT A CU¹ LZ / 00

Qy	238	MEDYKKNKKTIEINELIB--BSKTIIDKNKNATKEEBKKLYOAOQVLDLSIYNKQLEEA	295
Db	782	NKDYNNONNONNIEN--NONIENNQNNONNIENNQNNONNONNONNONNONNONNONNQN	840
Qy	296	HNLSV-----LEKRI-----DTLKKNEI-----	318
Db	841	NAGHISPDFFLYKMSLEKFIQFTKEHIVYIMDEKKKTSNIKESISLKKHHTKSI	900
Qy	319	LDKINETKNPPANGTNPNTLID-----KNNKIEBEKEIKE-----	356
Db	901	INNNDNN-----NEDDDNDMLSVMSYNDVDYIKNR--KENNKEILEEHVSFSSKNNEY	954
Qy	357	IAKTIK-----FNIDSL-----FTDPLELEYLYLRKNKNIDISA	390
Db	955	IAHSFNSTILLQLSNYIENIECKQMEYFKNSLLSYVDDYTITIELFYRLKKYNNIFSTEQ	1014
Qy	391	KVETKESTEPNEYPNGVYTP-----SYNDINNALNE-----LNSFGDLINPFDTYK	437
Db	1015	ILGTQYPSILQKLHDGI--YCLDKNNKKKNNDGNSQEDDDGKNKKNDGNKSOEQDDG	1073
Qy	438	EPSNIYTD-----NERKKFIEIKEIKIEK-----KKIESDKSYEDRSKS	480
Db	1074	NKKKNNGDNKSQEDDYGNNKKKNDDDDSYKIELIVDELNCKKKNYTDIELYELMKGS	1133
Qy	481	LNDITKEY-----EKLLEIYDSKFN-----NNIDLT--	507
Db	1134	DFDIFKKYKILNHFNIINNIFSTTISPHIHNIEDKYKVVYERFNLFNFNFSNVELSF	1193
Qy	508	-----NFEKMMGRYSKYVEKLELPHHTTFASYENSKNLEK-----LPKALK-----	555
Db	1194	DLIRREDKIL--RLTKYQOLLEEN-----YEKIKNKEKEEYLHACIKELMNLERYNN	1247
Qy	556	RNIYV-----EKELKYKKNLISK-----IENE-----TETIVENKKDEEOLPEKK	596
Db	1248	EKIILDOINEKEKKI--NIINEKYILLEKEEYEQNNKIFNAQIENLEKEKKOLQBEI	1305
Qy	597	ITKD--ENKPDEKILEYSDIVKY--QVQKVLNMMKIDELKKTOLIK-----NVELK	644
Db	1306	IQKMINVKLEK--NCDIIKIYEKEKOVLYHTLQENKDSHNYLKKDPENLLNLEKLL	1362
Qy	645	--HNHVPN-----SYQBNKQBPYYLIVLKKEDKLVKVPKPYBSLINEKK-----	690
Db	1363	YDHDISLKNKINTLWLEKEKNKKNTFHMNLRYENNNNLLJMKKEQONRYIKKELNERI	1422
Qy	691	-----NKTEQSDNSEPSTEGET-----TGQATTKPGQOAGSAL-----	725
Db	1423	KQINVRNNTSLSLRDN--STRGSIHQINNMNMNTHLPGMGASKINNINSIYNSNM	1480
Qy	726	-----EGDSVQAAQEQKQAOQ-----	742
Db	1481	IHMHSRSI IKNKEDAGNSTQARMNKDSTONI INNIHNTDINNMMNNINNTLNSINS	1540
Qy	743	-----PVP-----VVPPEAKAQVTPPAPVNNKTEVNSKLDYLEKLYEPLNTSY-----	786
Db	1541	NHLYPPFPFHNNVNSPKWGVGCDVTLASGVNK-----KDDFLNLNEENBENSFLYEYTRI	1595
Qy	787	-----TC-----HKYITLVS-----HSTMNEKI-----LKQYK-----	808
Db	1596	KSJAEELCKRESEILKTKGKNKILITCIETWKCFCKNSKEEISRLKEICKOLEKHKFEL	1655
Qy	809	--ITKEESKLS-----SCDPLD-----LLFNIGNNIPVMYSMPDLSNNLSOLFMEIYEKE	858
Db	1656	LINKSNEDKLKYINSLACDEKDYDIVVKDIKN--MRNEIDKLN-----DINEKS	1705
Qy	859	MVCNLYLKLDNDKTK--NLLLEEAKKVSTSVKTLSSSSMQPLSLTPODKPEVSANDDTSHS	916
Db	1706	YEIKLLKHENNLINEMNLKNETENMNKIQ-----KEEDYIKLIKDKDTNIQ	1754
Qy	917	TNLNLSLKLFENILSLGKNKIYOE--LIGOKSSENY-----EKILKSDSTFYENES--	967
Db	1755	NEYNDLLEKNEVVVV--KNNNLTYNDMNVLLKHEEIFLLKENIKITQKDNWYILNMFKN	1812
Qy	968	-----TNFVKSADDIINSLNDESKRKKLEEDINKLKTCL--QLSFDLYN--KYKLELRF	1019

Db 1813 QINYVDNNLLKRLDQLFNINQD-----LQKHLDTNOKHLEQLKYDIEIKERLAKIEKTK 1867
QY 1020 DKKKTVGKYMOKIK-LTLKLEQESKLSLNNPKHVLON-----ESVFPN 1064
Db 1868 INKQE-KYIIQLOKDNILDFNSTTTTTNNNNNNNNNNNNNNNDTVOQFTH 1925
QY 1065 KKKEABIAETENTLENTKILLKHYGLVKYNGESSPLKTLSEESIQTEDNYASLENPKV 1124
Db 1926 SLKA-----NLNSRLLEKLSNL-----NEKQLSDE-----KNRMK 1958
QY 1125 LSKLEGLKDNLNLEKKKLSYLSGLHLLIAELKEVI--KNKNY-----TGNPSPE 1173
Db 1959 ITILEDKLFKN--EKDKM-----KLOQIIDNNKNYMIQYNKLTNLDMLSE 2003
QY 1174 NNTDVNNALSYKKFLPE-----GTDVATVSESGSDTLEQ-----1209
Db 2004 ENRMILLNKEEYEQIEQLNHDHKLFIETKNNDIQIENKELOEQVDQYITINEKDKII 2063
QY 1210 ----SOPKKA--STHVGAESNTITTSQNVDDVDVIV--PIFGESEEDYDGLGVVY 1261
Db 2064 VHLNLQIKLANQNEHMRSCDIFNVASHQDNIKNHMHVVGEDIMGDTNHDVKNKIDQGT 2123
QY 1262 GEAVTPSVIDNLSKIENEYEV-----YKPLAGVYRSLKKOLENVV-----WTFN 1308
Db 2124 NOHINQGTNOHI-NOGTNOHDTCDGPNYNYVKVQVQATNREDNKNERNLSQEIYRYINEN 2182
QY 1309 VNVKDIILNREKRENFKNVL-ESDLIPYK-----DLTSSN-----YVVKDPYKFL-- 1353
Db 2183 IDLTSELEKKNMLENKNELEKEEYKLLNNDMLSNCKKLESIMMEKYKILNN 2242
QY 1354 -----NK--EKRDKPLSSVNYIKDSI-----DTDI--NFANDVLGYKILS 1390
Db 2243 NNIQEKDEIENLKNYKNNKLDLNNYSVVDKSIWSCPDSNINSPSCNDILNVFNLS 2302
QY 1391 EYK-----SDLSIKYIYNDKQENKYLPLFNIIETLYKTYNDK-----IDLEV 1436
Db 2303 KSNKKVCTNMDCNMDSISSNNYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2362
QY 1437 IHLEAKVLNYYTEKSNVEYKIKELNYLTIQDKLADFKKNNNFVGIADLSTDYNNH-- 1492
Db 2363 VN-----NLQNKNDNIIIIKFNILKL-----FKLGSCLYI-----INRNLEI 2403
QY 1493 -NLLTKFLSTGMVFENLAKTVLSNLDGMLQGLMNLISQHCYKQCPQNSGCFRHLDERE 1551
Db 2404 QMLKQILSLESIKSL-NEFINLNKEN-----EKNELIK-----INNPE 2443
QY 1552 ECKCLLNYKQEGDKCVENPNTCNE 1577
Db 2444 EILKLNQLQDNESCIONLNNYLKKN 2469

RESULT 29

G70163
hypothetical protein BB0512 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70163
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID: 98065943; PMID: 9403685
A:Accession: G70163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2166 <KLE>
A:Cross-references: GB:AE001153; GB:AE000783; NID:g2688419; PIDN:AAC66876.1; PID:g268842
A:Experimental source: strain B31

Query Match 6.0%; Score 507; DB 2; Length 2166;

Best Local Similarity 20.8%; Pred. No. 1.6e-08;
Matches 355; Conservative 306; Mismatches 552; Indels 494; Gaps 88;
QY 132 VRNYLLTTELKELYPQLFDLTJNHMLTCLDNHGF-----KYLIDGVEETNELLYKLN 182
Db 9 VNLFLSVILFVYRQ-YDKRSRAL---DKIKKFPVDLTUKVNLDFIEDTKTEINDLAVDME 64
QY 183 FY-----FDLLRAKLNDV-----CANDYQOI-----PFWLKIIRAN----- 212
Db 65 AVORSSEIIEK-KIEVQOKIKNKSNDFAVEKKIAYHDSMLKOLDDEMFKVQENIQLQ 123
QY 213 ----ELDVLUKLVFGYRKPDLNKNVKG-MEDYIKNNKKTIE-----NINEL 255
Db 124 VDGKIVDKLSKTLKGFNTQIDSVESNLVLEKFKDRANKENLESIKIASWEKFDTKIKEL 183
QY 256 I-----ESKKTIDKNKN--ATKEEEKK-----LYAOQYDLSLYNNKOLE 294
Db 184 VFKIDNLNKEISLYEKDLANIEERKNDILVKGNEKLDLDFSEFLEKVEFNIGYKSEIES 243
QY 295 AHN-----LISVLEKRIDTLKKNENIKE--LLDKINEIKNPPPPANSNGTPTNLL--- 341
Db 244 SFNFEYENKYKLIENSTELTIMESVKNKINEKEDFILNRLNE-----ELQNKFKDILVY 296
QY 342 -DKNKKIEBH-----EKEIKETAKTKFN-----IDSLFT 370
Db 297 DDRSKEIKDKLEDKLVLDVNEISSSSSPKDNVYSRINSLEESMRITEMGKYEQVDVDF- 355
QY 371 DPLELEYLREKKNKIDISAKVETKESTEPNCPVTPSYNDINNALNELNSFGDLI 430
Db 356 DKFRSQVLENLKN-----IYEDYEDKISQVDNNIRERV--ELSLDLNLSKMSVQS----- 404
QY 431 NPFDYTK---EPSKNIYTDNERK-----KFTNEIKE-KIKIEKKIESDKKSYED 476
Db 405 GAIDFTKRLDDSDNGIYLEFKGKFGADIEVFSSEFKGIDINQLKMOLESQLLDVSNIQEK 464
QY 477 RKSJLNDITKEKELNLIYDSKFNNDILDTNPEKMMGRKRYKYVEKLTHTHTFASYENS 536
Db 465 LIKLDNLLSNFEEI-----NGRFNNN--YSNLDNINAKYTALFESL--DSSSSKFEQ 515
QY 537 ----KHNEKLTGK-----ALKYMEDYSURNIWVEKELYYKNLISKIENETLTVENIK 586
Db 516 MESKYSFTDKLTAGNDEFSLMYGEKFPQLSQEATNNYQEFQDLNKKLENEIESFNMEPE 575
QY 587 KDEBOL---FEKIKTDENKPKDEKILEVSDIVKVOVKVLLMKNKIDELK-----KTO 635
Db 576 KTOETLKVDFNTSLINIKDEIGKNIVEFRDYYDEVN--IFVTQLEESKLQYSKMOGEMD 633
QY 636 LILKNVELKHNTHVPNSYKQENKQEPYILVLKKEIDKLKVPMPKVESLINE-----EKK 690
Db 634 SNLKNIIE-----SQINKTNEEFSLIQLQDKK---GIELSSEFNDLSHIQKK 679
QY 691 NIKTEGSDNS-----EPSTEGEITGOATTK---PGQOAGSALE-----GDSV 730
Db 680 AIDMHGSKWDELIALNKSLLDIKVSSE-ELLSATLIKLESLEKVDNDRMEYVLLKTDIE 738
QY 731 QAQAEQKQAPQPPVPPVPEAKAQPTTPAPVNNKTE-----NVSKLDYLEKLEYFLNPTS 785
Db 739 SLVIEKYKELK--DMSYSQSDAEILGKEFINRQETFEIKDKSVFMLEDLNKKFKDDKN-N 794
QY 786 YICHK-----YILVSHSTWNKILKOYKTKTEE--ESKLSSCDPL-----DLLFNI 829
Db 795 FVISEEDCYKLKDKFKIESEDILNFKSDLNEFIESKLQIVSNIKSDNQKQIDDDLDRI 854
QY 830 QNNIPVMVSMFSLNLSLSQLFWEIYKEWCVNLYKLKOND---KIKNLEEAKEAK----- 880
Db 855 SKDI---LNRKSIINNEVDKLSLSD-----WOSKLNIEITVKIENULLSSGKVDLDLI 901
QY 881 --KVSTSVKTL---SSSSMQPLSITPODK-----PEVSANDD 912
Db 902 DSEVTTIKELKFSLESLESYILEKIDEFNRQAGIYSDQLQIMNHNFKETRELENLS 961
QY 913 TSHSTNLNLSKLFENLSLIGKNK-----NIYQELIGQKSSSENFYEKILKDSOT 961

Db 241 KKIAELVQAQKEIAKTIKFNLDGIFVDPVLEEF 274

RESULT 32

S67593

transport protein USO1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein D2552; protein YD1058W

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000

C:Accession: S67593; A38455; S30782

R:Bloecker, H.; Brandt, P.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67587

A:Accession: S67593

A:Molecule type: DNA

A:Residues: 1-1790 <BL0>

A:Cross-references: EMBL:274106; NID:g1431058; PID:g253003; PID:g1431059; M

A:Experimental source: strain S288C

R:Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.

J. Cell Biol. 113, 245-260, 1991

A:Title: A cytoskeleton-related gene, USO1, is required for intracellular p

A:Reference number: A38455; MUID:91185402; PMID:2010462

A:Accession: A38455

A:Molecule type: DNA

A:Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>

A:Cross-references: GB:X54378; NID:g4777; PIDN:CAA38253.1; PID:g4778

A:Note: the authors translated the codon ACT for residue 768 as Ile

R:Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kend

submitted to the EMBL Data Library, February 1993

A:Description: An integrin analogue in Saccharomyces cerevisiae.

A:Reference number: S30782

A: Molecule type: DNA
 A: Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S'
 A: Cross-references: EMBL:L03188
 C: Genetics:
 A: Gene: SGD:USO1; INT1
 A: Cross-references: SGD:S0002216; MIPS:YDL058w
 A: Map position: 4L
 C: Keywords: coiled coil; transmembrane protein
 F: 326-342/Domain: transmembrane #status predicted <TM1>
 F: 334-410/Domain: transmembrane #status predicted <TM2>
 F: 617-633/Domain: transmembrane #status predicted <TM3>
 Query Match 5.8%; Score 487; DB 2; Length 1790;
 Best Local Similarity 20.6%; Pred. No. 5.4e-08;
 Matches 314; Conservative 278; Mismatches 561; Indels 368; Gaps
 Qy 98 GSVAAGGSGNSRNTNPDSNDSSDASKSYADLKHRYRNYLLTIKELKYPQLFDLTHMLTL 157

438	Qy	DN	AKHGGSNKSUKSDSGDFDGDGTEI	EGSGKANDFEVL	YUUDLUNLAPFALUF	314
158	Qy	CD	NTHGKFKYLIDG--YEEINEL	-----LYKLNFYDILLRAKLN--DV---	195	
515	Db	TT	DIEMFFQDHHKYSSELRREITRNVT	TGNDLEDEEPLKAIQTISELLTTS	574	
196	Qy	-----	CANDYCO	-----IPPNLKRANELD-V--LKLKLVFGYRPLD	229	
575	Db	IS	YLFYLIYLFDFGPKATNDFV	LSDKSVIKSLFSFYQIDEDVTITKCLVTLMLVAYEFS	634	
230	Qy	NK	DNVGMEDYTKKKNTI--ENINELLE	ESKK-----TIDKNKNA--TKEEK--	277	
635	Db	S--	KESPPEYEFYTKLTKDNVYASRK	QFKQKDSYFSKYDMDNDSILTPELDETGLPK	693	
278	Qy	LY	QAYDLSYINKOLBEAHNLISVLEK	RDTLKKNKENTKELDKINETKNPPPSANGNTP	337	
694	Db	VY	ESTYFIQLFNENI	-----YRIRTSALSHDPOEPIKNTS	728	
338	Qy	N	LLDKNKKLTTEEHEKEIKETAKTI	KFNIDSLTFDPLELEYLYREK-----	392	
729	Db	-----	FEVEKLRQCTK--LKGEITSLQ	TETESTHENITKLIATNHEKLEDEKY	778	
393	Qy	ET	KESTEPNYPNGVYTPISYNDINNAL	NELNSFGDLINPFYDTKEPSKNITYTONERKKF	452	

Db 779 QILNSSHSLKENFISILETELKKNVRSLSDEMTQLRDVLETKD--KENQTALL---EYKST 833
QY 453 INEIKKIEKKIE---SDKSYED---RSKSLNDITKE---YEKLLNEIYDSKFN 502
Db 834 IHKOEDSIKLEKLETLISQKKAEDGINKMGKDLFALSREMQAVEENCKNLOKEDKS 893
QY 503 NIDLTFKMMGRYKVKYKLEKTHHTTFASYENSKHNLEKLTAKALYMEDYSLRNIVVEK 562
Db 894 NVNHQKETSLSKEDIAAKITEIKAIN--ENLEEMKLOCNLSKEKEH-----ISK 941
QY 563 ELKYK-----NLJSKIENETIETVENIKKDEQOLFEEKITKDNKDPKILEVSDIV 615
Db 942 ELVEYKSRFQSHNDLVAKLTLEKLSLANNY-KDMAQENESLIKAVEESKNESSTIQLSN-- 998
QY 616 KVQVQVLLMKIDELKKTQILKLVNKLHNIHVPSYQENKQEPYLLVLKKEIDK 675
Db 999 -----LQNKIDMSQE-----KENFOIERG-----SIEKNLEQLK 1028
QY 676 VFMKPVESLINEKKNIKTGQSDNSEPSTEGEITQATTKPQQAAGSALGDSVQAQAQ 735
Db 1029 ----KTISDLEQKELIS--KSDSSKDEYESQI---SLLKLEKLETATTANDENVNKISE 1079
QY 736 EQQAQPPVPVPEAKAQVPTTAPVNNKNTENVSKLDVLEKLYEFNLTSYICHKILVS 795
Db 1080 LTKTRE-----ELEAEL---AAYKNLKNLETKLETSEKALKEVKE----- 1117
QY 796 HSTMNKILKOYKITKEESKLSCDPLDLFNION-----NIPWYSMFD----- 841
Db 1118 ----NEHLKEEKIQLEKATETATKQOLNSRLANLESLEKEHEDLAQLKYEQIANKER 1173
QY 842 SLNLSQLFMEIYEMVCNLYKLDNDKIKNLLPEAKRVSTSVKTLSSSSMOPISLTP 901
Db 1174 QYNEEISQLNDEITSQOENESIK-KKNDLEGEVKAAMKSTSEQSNLKKSEIDALNL-- 1230
QY 902 QDRPEVSANDTSHSNL-----NNSLKL-----FENILSLGNKNI 938
Db 1231 QIK-ELKKNKNETNEASLLESIKSVESETVKIKELQDECNFKEVSELEDKLASEDKNS 1289
QY 939 -YOELIGOKSENFEKILKSDST---FYNESFTNFVKSADDSLN-----DESKR 988
Db 1290 KYLEL--QKSEKIKEL--DAKTELKIQLEKITNLAKAKESSELSPLKTSSEERK 1345
QY 989 KLEEDINKKTKTQSLFDLYNKYKLERLFDKKTGVGYKMOIKLTLKEQLESKLS 1048
Db 1346 NABEQLEKLNKEIQIK-----NOAFEXER-----KLLNEGSSTIQEYSEKINT 1389
QY 1049 LNNPKHVLONFVFNKKKEAEATAENTLENTKI---LLKHKGVLKYNGESSPLK- 1103
Db 1390 LEDELIRLQN-----ENELKAKEIDNTRSELEKVSNDLELEEKONTIKSLQDEILSYKD 1445
QY 1104 --TLSEESIQT--EDNYASILENFVLSKLBGLKLDNLNLEKKLSYLSGLHHLIAELKE 1159
Db 1446 KIRNDEKLLSIERDNKROLESLEKQOLRAAQESKAKVEEGLKLEESSEKAELEKSKE 1505
QY 1160 VIRKNKVTGNPSNENTVDVNNALLESYKFLPEGTDTVATVVVSEGSDTLEQSQPKKASTH 1219
Db 1506 MMKLESTIES---NETELKSSMETIRK-----SDEKLEQS--KKSAAED 1545
QY 1220 VGAESNTITTSQNVDDVDDVIIVPFGSEEDYDDLQGVVTEAVTPSVIDNLSKIEN 1279
Db 1546 I-----KNLOHEKSD--LISRINSEKIDIEELKSLKRIEAKSGSELETVTKQELNN 1593
QY 1280 EYEVLYL-----KPLAGVYRSLK-KOLENNVMTENVVNDIILNSRKNREKNVL 1329
Db 1594 AQEKIRINAEENTVLSKLEIDIBRELKDKQAE---TKSNQEEKELTSLRLEKE----- 1644
QY 1330 ESDLIPYKDLTSSNYVVKDPYKFLNKKRDKKFLSSNYIKDSIDTDFINFA-----NDVLGY 1385
Db 1645 -----QELDSYQQAOK-----SEERRAEVRKQVEKSQOLDKAMLETYNDLVN- 1691
QY 1386 KYLISEKYSDDLISIKYINDKOGENEKYLPLFNLTIE---TLYKTVN-----DKIDL 1436

Db 1692 ---KEQAWKRDETVKTKTTDSQRQEIEKAKELDNLKAENSKLKEANEDRSIDDLMLLV 1748
QY 1437 IHLEAKVNLVITYEKSNVVEVKI 1457
Db 1749 TDLDERNAKYRSKLKDLGVEI 1769
RESULT 33
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
J.Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
R.Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156
Query Match 5.7%; Score 483; DB 2; Length 1127;
Best Local Similarity 19.7%; Pred. No. 4.2e-08;
Matches 287; Conservative 247; Mismatches 430; Indels 490; Gaps 65;

QY 132 VRNYLTIRELKY-PQLFDLTNHLMLCDNHGFKYLDGYEELNELLYKLNFFDILLRA 190
Db 54 IKNNYKIVEKIFMYHMQFKINDNIIL-----QYLLIYNNNEINKCIENKF----- 99
QY 191 KLDVCANDYQCPFNLIKIRANELDVL---KKLVFGYRKPDLNIDKNVGMEDYIKKKK 246
Db 100 ----PCKNPLYNTYKKKLYIDLYEKKDKELVINIQ-----KNADVIND-IGNV 149
QY 247 KTIENTINELIEESKKT-----IDKNKNATKEEKKLYQAQYDLSIYNKQLEEAHNLSVL 302
Db 150 NNTHSDNETIITGKETLIDILNKLKLVSSDEKQLEIQYKNNIN--NKEIE----- 198
QY 303 EKIDILTKNENIKELLDKINEIKNPPANGNTPTNLDKNKIEHEKEIEIAKTIK 362
Db 199 FRNIDNVQREINKQ--DELN-----LLDESKK--EFIKKOEELNKTID 239
QY 363 FNIDSLSF--TDPLELEYLREKNKID-ISAQVETKESTEPNEYPNGV--TYPLSYNDIN 417
Db 240 KQOEELIKKLNDEINFNIDKQKLLDQINSKINTL-----NENIKGVMLYETKKNIS 294
QY 418 NALNELNSFGDLINPFDYTKEPSKNIYTDNERKKFNEIKEKIKIEKKIESDKKSYEDR 477
Db 295 NLQNEILNKDSTIKSLD-----EKOKLDEL----- 320
QY 478 SKSLNDITKEVELLNEIYDSKENNNIDLTFEKMGMKRYSKYKVELTHHTNTFASYENSK 537
Db 321 DKNNITLSYLNK-----SNTKIINIQQLL-----BSSLTDFNNAN 356
QY 538 HNLEKLTALKYMEDYSLRNIVVVEKELKYKLNLSKIENTEI---ETLVENIKKDEEOLF 594
Db 357 ININEL-----KSKIKLFDNDIQKLNNDITEQNNKITDFFENNSTRIFK 399
QY 595 KTKTKDENKPKDEKILEVSDIVKVOQVLLMKNIDELKKTQILKLVNKLHNIHVPSYK 654
Db 400 EKLDTE-----YKKIDIDIKNNN--LQKLE-----ESYK 425
QY 655 QENKQEPYLLVLKKEIDKLVFMPKVESLINEEKKNIKTGQSDNSEPSTEGEITQAT 714
Db 426 KIDEQTEYY-----KNKINKEYNDI----- 445
QY 715 TKPQQAGSALGDSVQAQAEQKQAQPPVPVPEAKAQVPTTAPVNNKNTENVSKLDY 774
Db 446 -----TELKNNLOKLEENKK-----IDEQTEY-----Y 470

QY 775 LEKLEEFLETSYICHKYLIVSHSTNMEK-IKQYKTK-EESKSLSCOPDLLEFNIONN 832
 Db 471 KMK--INKEV-----NDILENNNNLQKLEENKKNIN---DKLTKLND 509
 QY 833 IPVMSFSDLSNNLSIQLFMEIYEREMCNLKKLNDKIKNLEEAQKYSTSVKTLSSS 892
 Db 510 IESNTELFNKLNIS-----DFKDKSR-----FIALNTIEYEDLRD 545
 QY 893 SMOPLSLTPQDKPEVSAANDTSHSTNLNLSLKLFNILISLGNKNKIYQELIGQKSSSENF 952
 Db 546 LLENINKT-----NELMKLSDNKLS-----SLEQLY 571
 QY 953 E--KILNDSDFYMESEFTNPFYKSKADDINSLNDESKRRKLEEDINKLKTITQLSPDLYN 1009
 Db 572 DSKNNILGIDIDYNS--LKEKND-----KIDYEFSNIEK-----FDIYN 609
 QY 1010 KYKTLERLFDFKKTVGKKMOIKRL--TLKLEQESKLSLNPNRKHVLQNPSEFNKK 1066
 Db 610 VI-----ENKIGLNDSITNKIINNDQKREYINSKIDSKSDELSTW--FDDIFNKK 658
 QY 1067 KEAEIETENTLENTKILKLHKYGVKKYVNGESSPLKTLSEESIOTEDYNVAS-LENFKYL 1125
 Db 659 N--QLASITNNIENI-----SNKIKDNLNFIINSDESKLELDGIRKY 699
 QY 1126 SKLEKLLKDNLNLKLEKKL-SYLSSELHLHLAEKVEYIKNNKYTGSPSPENNTDVNNALES 1184
 Db 700 KOOFDKIKDAMTEVKSSEENTLQDKIDSISKINSINELTNAYDIINTKANDLIDKLNNYGSE 759
 QY 1185 YKFFLPEGDVATVYVSEGSDFLEQSOPRKPASTHYGAESNTITTSQNVDDDEVDIYIP 1244
 Db 760 FKNLYNNAASDLL-----DTIQNNDKRYKQMLEYELEKN-NOSIEINDIVNNFIKEL 810
 QY 1245 IFGESEEDYDDLGOVYVTEGAVTPSYI-----DNILSKINEVE-----VLYLK 1287
 Db 811 IKFNNTETNKSINELLTJMDINDIKLFLKLYKELNKTISTNNLKITYNEINDVNAEKISYIE 870
 QY 1288 PLAGYRSLKQLENNVTFNPNV-----VKDILN-----SEFNKRENF 1326
 Db 871 NLQFINSFSLSEFNQGSITSHINFLNLTLAGINDVNLKNTKIMADTTRRGDGTNRDEIK 930
 QY 1327 NVLESDDLPRYKDLTSSNVVVKPYKFLN-KEKRDKFLSSYNTIKSDIDPIDFANDVGLY 1385
 Db 931 NQISEENIKSQKFNKN--EKDCLKLISFNDKLNTYNISAGYTEYN-----NIEHECLKY 983
 QY 1386 -YKILSEK-----YKSDLSIKKRYINDQGENEKYLPFLNINTELTLYKTVNDKIDLEFVHLE 1440
 Db 984 LYIAVSDQEVRYKFFIHNKIQIDKNKKNSMQFLKLF-----LKRITIIYD 1030
 QY 1441 AKVLNITYEKSQVNEKIKELNLTAKTIQDKLADFKNKNNNFVGIADISTDYNNH---NLIT 1496
 Db 1031 LKILSDTID--NIQKQYLLNKLKYNEKFKLNTFEO-----FINQYGTGFMPIFNILN 1080
 QY 1497 KFLSTGVFENLAK 1510
 Db 1081 --ISTNMVFYIFK 1092

RESULT 34
 A42771
 reticulocyte-binding protein 1 - Plasmodium vivax
 C:Species: Plasmodium vivax
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
 C:Accession: A42771
 R:Galins, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
 Cell 69, 1213-1226, 1992
 A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites
 A:Reference number: A42771; MUID:92315336; PMID:1617731
 A:Accession: A42771
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1:2829 <GAL>
 A:Experimental source: Belen strain, merozoites
 A:Note: Sequence extracted from NCBI backbone (NCBIN:108114, NCBIP:108115)

[illegible]

RESULT 39
B42771
reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
C:Species: Plasmodium vivax
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 09-Sep-1997
C:Accession: B42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites
A:Reference number: A42771; MUID:92315338; PMID:1617731
A:Accession: B42771
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1252 <GAL>
A:Cross-references: GB:M88098; NID:g160627; PID:g160628
A:Experimental source: strain Belem, merozoites
C:Genetics:
A:Gene: RBP2

Query Match	5.4%	Score 457.5	DB. 2	Length 1252
Best Local Similarity	21.1%	Pred. No. 2.9e-07		
Matches 306	Conservative 228	Mismatches 458	Indels 455	Gaps 71
OY	272	EEKKRIQAOYDLSYVKOLEEAHNLISVEKRIIDITLKKKENIKELDKINEINPEPPA	331	
Db	1	EFDEKKEVADTSFDEK-KKSIEKAY-----EKMGNLTKE-----LEKMDKEN----	41	
OY	332	NSGNTPNLLDKNNKIEHEHEKEIKELIATIKFNID---SLFTDPLELEYLREK----	NK 384	
Db	42	-----IEKEVEEAQIQYKRIF-----IDHDVNLMDVEYKSKIVAEKITYELYYK	84	
OY	385	NIDISAKVETKESTPEPNYNGV-----YPLSYNDINNALNDSFGDLINPFDTKEPS	440	
Db	85	EID-----EIQKLT--NEYKOGDTSNFFYTEQYNSATGSKAKIEQFINATTKKTSPTS	137	
OY	441	KNITYDNERKKFIEIKERIKIEKKKIEBDSKSYEDRSK---SLNDITKEYEKLNN--EI	495	
Db	138	ODI---NE-----LESIEKEVHKNLQYKQESMSMEEMKQILSMKDL-----ILNNSET	185	
OY	496	YDSKPFNNNIDLTNPEKMMGKRYSYVEKLTHTHNTFASYSNENHNEKLTKALKIYEDYSL	555	
Db	186	IAKELISNNT-----ONALGFREMAATKTKMKIDELLQ-----	216	
OY	556	KNIVEKELKAYYKNLISKIENIEIETLVENIKKDEQLFEKKITTKDENKRPDEKILEVSDIV	615	
Db	217	RVAAMIEEAKHKK-----NNIDIALEDAQID-----TEVSKIE	249	
OY	616	KVOYQOKVLLAMKIDELKATQOLLKKNVLEKHNHIVPNSTYKOEKQEPYLLVYLKEIDIKR	675	
Db	250	QINRR---INMKKEIKESYLSIEIKYKPKCTTELSNKRGRDK-----IEELE	294	

QY	676	VFMKVES-----LINEKNKNTIEGO-----SDNSPSTGECEITGATTPGQOASA	724
Db	295	KFKPBEESNSKNVNIENINRNSQYLDKIEDAKQASTVEFLPHKHET--TITSINF	351
QY	725	LEGSVOAOEOQOAPRPVPPVPEAKQVPPRPAPVNNKTEVNSKLDYLEKELEPANT	784
Db	352	KESLEILOYETKSQKKI-----NKADIMK--EIER-----	379
QY	785	SYCHKAYILVSHSTWNEKIIKQYKTKREESKISSCDPLDLFNIONNIPMY--SMFSL	843
Db	380	-----HNSEIOTQVKGQOENIKNL-----NPNHYDAEDEL	411
QY	844	NNSLSQFMEIYEKEWVCNLYKLKDN-DKJKNLEAKKVTSTVKTLSSSQPLSLPO	902
Db	412	NMDKSTNAKYLE-----TNLESVKHNLSETITNKQGEKITYSKADI-----MOKKATSE	463
QY	903	DKPEVS--ANDDTSHSTNLSNLSKLEFNILSGKNK-----NIYOLIGKSSSEN	950
Db	464	NTAETKTELEKVDQSDSNVYNLNOITERNLIVETEKNRJNGDISTTINIGAL---KESKG	520
QY	951	FYE-----KILDSOTFYNESTNTNYSKADINSLN--ESKR	987
Db	521	NTEIGFLEKEIEICKNKRLKVDITKKSINSTGNFSSLTNNFDLNOYDKNKMINIYENKM	580
QY	988	----KKLEEDINKKLKTLQLSFDL---YNKYK-LKSE-----RLFDKKKYTGAKMOIK	1033
Db	581	GEIYNFEGSLNKISEMLNRASENTSDYNSAKTLRLAEQKEKYNLNNKEEANKTLROYK	640
QY	1034	KLTL-----LKOELESKLNSL-----NNPKAYLONFSYFNKKKKEAEIAE--	1073
Db	641	KVESREFLENNKESLD-KIEMIMIKBQLTVNEGHGVKQJVAENIKELVDENNLSDILKQA	699
QY	1074	-----TENPLEN-TKILLKHYKGLVYKYNNGASSPPLKTLSEESIOETEDNYASLENF	1122
Db	700	TGNKEEIOKITHSLKAKKATIIIGHDYTSAKYGIKITTELALTE-----	744
QY	1123	KVLSLECKLKDNLNLEKKKLSYLSGLHLHLAEKLEKVINKNYTGNSPSENNTDVN-NA	1181
Db	745	LILDADAKLTAQAEIKFEFSK-----NNVULETEMNSKN-----TNELDVHKNI	785
QY	1182	LESYK--KFLPECTDVAATVVSSESG-----SDPLESOOPKPA--ST	1218
Db	786	QDAKRAVALEILASHDEIDTYOKOSSKILIEKMGNOIYLKVLINQYKKNKSIKSKEANSY	845
QY	1219	HVG-----AESNTTTSQNVDEVDVYIIVPIFGESEBDYDGLQOVYTGAEVPSVIDN	1272
Db	846	KIGVNSKSHSELKSTICS--DKSYNTIIAL--EKOTELONLNSPFOEK-TWTNDS	897
QY	1273	ILSKIEHNEYVL--YKLPLAGVRSILKQOLEN--NMFTFNVAVKOLNSRFRKREPNFYV	1328
Db	898	KLEKIKIDFESLKNALKTTEGVEYNALKASDNHEHQSSEPNPAL-SELEKET----	952
QY	1329	LESDLIFPKDITSSNYYVKQPYKFLKKEKRDGFILSYNYIKDSI-----PTDI--NPNAN	1380
Db	1000	NKAYKYLAVIKKNVEDYQODVLTLNEHFNTKQVSNHEPTNFDKSNKSSEBELTAAVYDSKTI	1059
QY	1429	NDKIDLFVILHEAKVUNYTYEKSNNVYKIKEL-----NYLKTIODLADKKNNNVGJAD	1484
Db	1060	ISKUKGYITIE--VNEENTEMTITSSAKELAEALYNELKNNKTSJNEIYOTSNEVKQDE	1114
QY	1485	LSTYUNNNNLTKFLSTGWFENIMAKTYVSNL-----LDGNLOGLMINSOCH-	1530
Db	1115	MKSNAAD-----KYIDVSKIFNTVLDTQKSNITVTNGHSHNNVKNKDLKGLDELIDADSSF	1166
QY	1531	--QCVRK 1535	
Db	1169	TLESIRK 1175	

A:Gene: GDB:CEPHE
 A:Cross-references: GDB:361164; OMIM:117143
 A:Map position: 4q24-4q25
 C:Superfamily: centromere protein E: kinesin motor domain homology
 C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
 F:7-335/Domain: kinesin motor domain homology <KNOT>
 F:86-93/Region: nucleotide-binding motif A (P-loop)
 F:486-2183/Domain: coiled coil #status predicted <COI>
 F:92/Binding site: ATP (Lys) #status predicted

Query Match 5.38; Score 443.5; DB 1; Length 2663;
 Best Local Similarity 18.48; Pred.No.1.9e-06;
 Matches 345; Conservative 349; Mismatches 617; Indels 559; Gaps 82;

126 ADKRRVRYNLTITIKEL---YPLFDLTNNHMLT-----LCDNHGFXY 166
 415 AKRRRTWGLGKINKNKSNVADQFNIPNITTKIKLSINLREIDESCSDVFSN 474
 167 LIDYEEL---NELLYKLNFFDL--LRKLANDVCANDYCOI---PFLNLR-ANE 213
 475 TLDTLSIEWNPAKTLNQNIESELNSLRADYDNLVL-DYEQLRTEKEEMELKREKD 533
 214 LDV-----LKKLVFG---YKRPIDN-----IKDNVVK 237
 534 LDEFEALERTKKDOEMQLIHEISNLKLVNHRVYMODLENELSKVEILLREKEDIK 593
 238 MEDYIKKKNKTIENIN-----ELIEESK-----TID-KNKNAATKEEEKKL 278
 594 LQEIYDSOK--LENIKMDLSYSLESTIEDPKOMQTLFPAETVALDARRESAFLRSENL 651
 279 YQAOYDLSYNNKOLEEHNLSVLEKRIIDPLKN--ENIKELDKTIEIKNPPRANGNT 336
 652 KEMKKEIATTYKOME---NDIQYQSOLKAKMQVDLEKELQSAFKEIYKLTSLIDGKV 708
 337 PNTLL-----DKNKRIEHE-----KIKELAKTK---- 362
 709 PKDLCLLEEGKITDQKELNKEVEENALREVYLLSELKSLPSEYERLKRKTQKSE 768
 363 -----FNIDSLTDPLELEY---YLREKNKNIDISAKVETKESTEPNXPNGVTPLS 412
 769 ELHITSEKDLSEVYVHKESRVOGLLEIGTKTKDLDATQSNVSKTDQEFONKKTLM 828
 413 YN-----DINNALNL-----NSPGDLINPDPY--TKESKNIYDNEKKF 452
 829 FEQKYKVALEENEMNOEIVNLSKQAKFDSIGALTELSYKQELQOETREVOEELNE 888
 453 INEIKERIK-----IEKKK-----IESDKSYEDRSKSLNDITKEYEKL 491
 889 MEQLKQLENRDSPLOQVEKEKTLITEKLOOTLEEVKTLTQEKDQLQLOESLOIERDQL 948
 492 LNEIYDSKFNNNIDLTNFEKMMGRYSYVEKLTNN--NTFAS-----YEN 535
 949 KSDIHDT-VNMMNIDTQ-----EQLRNALLESKHOETINTLSKISEEVSRLNMEEN 1000
 536 ---SKNLEKLTAKKMEYSLRNT-----VEKELKYTNLSKINLETLETL 581
 1001 TGEFKPEFOQKMWGIDKKQLEAKNTQTLTADYKDNELIIOQRKIF-SLQOE-KNEIQOM 1058
 582 VENIKKDEOLFEEKITKDEN-----KPDKEILEVSDIVVOYQVLLNMRKIDELKKTQL 636
 1059 LESYIAKEBOL---KTDLKNIETKTENQELRLLSDELKKQ--OEIYAQOEKNNAIKKEGE 1114
 637 ILKN---VLLKNIHVAPNSYKOEKNOEPYLLIYLKKEIKLAKFMPKVSLEINEEK--- 689
 1115 LSKTCRLALAEVEKILKSSQOLOEQOQ---LNVQSEMSQOKKINIEINLKNELKNKE 1171
 690 ---KNIKTE---GQSDNSEPSTEGELTGOATTKPGQOAGSALAGD-----SVQAOAQ 735
 1172 LTLEHMETERLELAQKLNENYEEVKSITKERKYLKELQKSPFERDHLRGVIRIEATGL 1221
 736 EQQAQRPVVPVPEAKAQVTPPAVNNKTEENVSKIDYLEKLYEFLNTSYICHKITYLS 795
 1232 QTKEELKIAIHILKHOETIDELRRSVSEKTAQIINTDULEK-----S 1274

QY 796 HSTWNEKILKQVITKEESKLSCDPLDLLFNIONNIPVYMSNFDLSNLS----- 846
 1275 HTKIQDEI-----PYLHEQGLLRNVKVKVSETQETMNLLELTQSTKDTTLARIEMER 1330
 QY 847 --ISQFMETIEYKEMCNLYKLDN-----DKTNLLEAKKAVSYSTYKTLSS 892
 1331 LRINEKFOE--SOEELSLTKERDNLKTIEALEVKHDKLEHIRE-----TLAKIOESO 1383
 QY 893 SMDPLSTPODKPEVANSNDQTSHTNLSNLSKLE-----NLSLGNKNYIOELI 943
 1384 SKQOQSLNMEK-----DNETTKIVSEMEQFKPDSALLREIEMLSKRL----- 1430
 QY 944 GOKSSENFYEKILKSDPTFN-----ESFTNFKSKADDINSUNDESKR-----KK 989
 1431 --QESHDEKMSVAKERDQLRQEVLOESDQKLENKEIVAKHLETEEBELKVAHCCKE 1488
 QY 990 LEEDINLK-----KTLOLSFDLYN-KYKLEERLPDK-----KTVGKYKMOIK 1033
 1489 QEEFINELRVNLSKETESTIQOLEAINDKQNKIOEIVKEEOLNKOISEVDEQNV 1548
 QY 1034 KLTLLKE-----QLESKLSLNPNKHYLVONFSVFENKKK-----EAEIATEN 1076
 1549 ELKQFKERKAKSALOSITSKMLELTNRLOESQEEIQIMAKEKEBKRYOEAQIERDQ 1608
 QY 1077 TLENTKILKHYGLVYKYNGESSPLKTSSESI-QTEDVYASLENFVLSKLGKLDN 1135
 1609 LKEETKEI-----VAKKESQOEKEYQFLKMTAVNEQOEKCEIENHLEKQEFQKLNEN 1662
 QY 1136 LNEKKSLSYSSGLHLLLEKLEVITKNKNYTGSPENNVDVNNALLESYKFLPEGTQV 1195
 1663 IETENIRLTQI--LHNLEEMRSVYKERD-----DLSVETLK--VEEDQL 1705
 QY 1196 ATVVSSEGSDELPOQPKPASTH-----VGAESNTITTSQNVDEVDV- 1240
 1706 KENAREITTDLEKQEBELIYVHMLKHOETIDKRGIVSEKNEISNMOKDLEHSDAL 1765
 QY 1241 -----ITVPIGSEEDYDDLGOVYVGEAVTPS--VIDNLSKIE----- 1278
 1766 KADQLKIOEBELRIAHMLKQOEETIDKRGIVSEKTDKLSNMOKDLENSNAKLOEKIOEL 1825
 QY 1279 --NEVEYLYLK-----PLAGVYRSLLKQLENNVMPF-----NVNKOILN----- 1316
 1826 KANHQILITLKQDVNETQKVSSEMOQLKQIKDQSLTSLALENLNLAQLEHNELEMK 1885
 QY 1317 SRFNKRRENFNVLESDLPKPKDLTSSNVVYKPYFLNKRERKFLSSYNYIKDSIDTQI 1376
 1886 SVMKERNLNRVEET-----LKERDQKES--LQETKARDL 1920
 QY 1377 NFANDVLGYTKILSEKYSKSDLSIKTYINDKOGENEKYLFPNNTIEFLTYTVNDKIDLEV 1436
 1921 EIQOE-LKTARMLSKHEKEFVNDKLRKISERTIO-----ISDIQDLKDSADELOKKI 1972
 QY 1437 IHLKAVLANTYKESNVEYIKELNYKTIQDKLADKKN-----NNFVGIAIDSTQY 1489
 1973 QEIQKKEQLQLRKVEDVNMHKKINEMOQLKQ--FEPYTLCKCEMDNQLTKKLHESL 2029
 QY 1490 NNNNLITK-----FLSTGVNENIAKTVLNS-----LIDGNLQGM 1524
 2030 EETRIVAKEDDELRIKESLKMEDQFIAT--LRMLIARORONQVYKPERKLLSDGOHL 2087
 2088 MESLRKCSRIKELKLR--YSMDHNYGLANRLSDLEKETEFRHMKKLYVLSYTV 2143
 QY 1560 --KQEGDKCV 1567
 2144 KIKEQHECI 2153
 RESULT 43
 B71612
 hypothetical protein PF0555C - malaria parasite (Plasmodium falciparum)

OY 1574 CENENGGCADACCTEEDSGSNGKRTG-----ECGRKPSYPLFDGTCSSSNFL 1623
DB 1857 DKNITINIKDDKNITNMKKNNKNNYLTILMNSQECG--FYYSIFNTLI--NDYNFL 1909

RESULT 44

A64224 hypothetical protein MG218 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999

C:Accession: A64224

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;

M.; Fumman, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.

, C.A.; Venter, J.C.

Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.

A:Reference number: A64200; MID:96026346; PMID:7569993

A:Accession: A64224

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1805 <TIGR>

A:Cross-references: GB:039699; GB:143967; NID:91045903; PID:91045905; TIGR:MG218

A:Experimental source: strain G-37

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: Mycoplasma genitalium hypothetical protein MG218

Query Match 5.2%; Score 440; DB 1; Length 1805;

Best Local Similarity 19.7%; Pred. No. 1.5e-06;

Matches 334; Conservative 301; Mismatches 566; Indels 492; Gaps 79;

OY 133 RNYLTT-----KELYPOLFDTLNMILCDNIH----GEKYL 167
DB 190 KNYLLANTIDLYNELDLENOKRLLSIEYENTYRSADNELQYVENIDQNOIOEFHQ 249
OY 168 IDGV--EELINELLYLNF-----YFDLLRAKLNDVCANDYCOQRPMLKTANFL--DYLK 218
DB 250 YQTRDELSDLERKIQTLKQELVDKESALRYKIDA-----DFYINAAELADYVAK 301
OY 219 KLVGYRRPRLDNI--KDNVGMEDYIKRKNKTENINELIE-----ESKKTIDKKNATKE 272
DB 302 QLSF-----QDGITKQNAQNHEDLVALNKEKDLNTOKEFMYLRQSALIDIK----- 351
OY 273 BEKKKLYAOYDLSYNNKOLEAHNLISVLEKRIIDLTKNE--NIKELDKINETKNPPA 331
DB 352 -----LQENELFAKHLNHQNEFE--OKOSDSLKLETETFKALQHNKEKN----- 397
OY 332 NSGNTPTWTLDDKNNKEHEKEIKETIKFNIDSLFTDPLELEYLREKNNKIDISAK 391
DB 398 ESATKSEELLNQERELEKRE-----IDTLTLO--ASLEYEHQRESSQLLDKQ 445
OY 392 VETKESTEPNEYRNGVYTPPLSYNDINNALNELNSFGDLINFDYTKPSKNITDNERK- 450
DB 446 NEVQOHQNEV-----AKKELDKERNLL-----DQCKV 475
OY 451 --KFINETKEKIKETKKIES--DKSYEDRSKSLNDITKEYEKLINLETYDSKFNANNID 505
DB 476 DSEALIFOLKREKVAQERKELELYLVKKQKODQKE--NELL--FEFKQKQ--HQADFEMELE 531
OY 506 LTNEKMMGRYSYKVKLTHNHTFAYSKNNLEKLTALAKMDEYSLRNIVYERELK 565
DB 532 AKQOELFEAK--HALER-----SFILDEKDLN--TAAQOLANES-----QLK 573
OY 566 YKRLNLSIKNETLTVENIKKDEQOLFEEKITKDEN-----KPEKILEVSD-- 613
DB 574 TDKSKSADPELMQNEKYNLQOEKQKLFQERTYPERNAAVLSNNLOQKRELLDQKELID 633
OY 614 -----TVQVQVLLMKNIDELKLTQILLKVNELKHNHVPVSY 653
DB 634 QLTGSPQOERLINQREKHELVASVEKQKELIGKLIQDFOSQTLASKNLAEREMAIKFE 693
OY 654 KQEKQOPYLLIYAKKRIKIDKLYMPVESL--INEKKNIKTEGQ-----SDNSEPS 704
DB 1714 -----LTKKVNQHN 1722

DB 694 KEIATERKOLL-----NDVNAEVIQADIQALQNSLQNERSELQAKORIAIDFNDLSKL 749
OY 705 TEGETGQATTPRQQAQSALESDGSVOAQA-----QEOKAQA--QPPVVPVPEA 751
DB 750 NEVELSLQKRLQELQTLQANQKQSYQNAQFEBEGLDKLNEKQAFNLRRKQKQMEVDAL 809
OY 752 KAQVTPPAPVNNKTEVSKIDYLEKYLEFNTSYICH-----KIYL 793
DB 810 KQRLSD-----KHQALNMQAEILDKTHE--LNNAFNLNHADQKSLQDLATVKEFKQLID 863
OY 794 VSHSTNMEK-----LKKYKTKRESESLSSD--PLDILFNQIN 831
DB 864 LERSALLEKQREFAENAVAGEKRWMSNKTSQLQKTYELTKQESQTOKELEKIAF--S 920
OY 832 NIPVWMSFSLNSLSQLEMEIYEKEMVNCYLKIKDNKIKNLNEAKKVSYSVKTLSS 891
DB 921 DLQGDYQVFELOKQ--QEFQIOEAKQ-----RELDKT-----AEK--NQGVKLELD 962
OY 892 SSMQPLSLTPQDKPEVSAN--DQTSHTNL-----NNSL-----KLEFNLISLG--K 934
DB 963 NRPQALQNOKQDVOAOLELEREHOHLNEQTAENQANESILKQREQLTKKIQAFHYELK 1032
OY 935 NKNIYQELIGQKSESENEYEKILKQSD-----TFYNESPTNFVKSQADINSND-- 983
DB 1023 KRNOFLALKGRFLAKEDQDQKQDELINWRKQFEKETDDEDAKKKELELEKIRSLSS 1082
OY 984 -----ESKRRKLEED--INKLKTQLQSLFDLYNKKYKLERLEDKKTIVGY--RMQI 1032
DB 1083 QSNVELEKREKRLATDFTNLKQVQHNQINRDQNS--QIRQFLERKKNQFRESNEANA 1139
OY 1033 KKLTLLEQLESKLSLNNKPRHVLQNSVFPNKKKEAETENT--LENTKILLKHY- 1088
DB 1140 KRAFLIK--RLRSFASNLKQLEALAIQLEFDDKDEQOKKELQOATLQLEQFEKKEKNFD 1198
OY 1089 --KGLY-----KYNGESSPLKTSSESI-----QTEENYASLE 1120
DB 1199 TEKORLVAIKTQCEKLSDEKAKALNOKLVELKNLSQTYLLANKKAQESQOQLOQKYNLL 1258
OY 1121 NFK-----VLSLEGKLDKLNLEKKRLSYLSGSLHLLAELKEVIAIKR 1164
DB 1259 DLKENLEKTLQDLKKHRSIFARLT--KFANDLREKKQLAKQRIYVDKNNRLKENERNL 1317
OY 1165 NYTGNSESENNTDVNNALLESKFLPBEGIVAVYVESGSLTLEQSPKKAASHVGAES 1224
DB 1318 HFLSNETERKRAVLEDDISYFEKQKQATD--AIIASH-----KEVKKKEGLQKLVEL 1370
OY 1225 NTFITSONVDEVDVLIIVPIFGESEEDYD-----DLQGVNTEAVTPSVIDNLSKI 1277
DB 1371 ETRKTKLND-----FAKFSQREFFENORLKLLEKLTQLOTOTNSNNFKTAIOEI 1422
OY 1278 ENEX-----EVLVLRP-----LAGVRSLLKQLENNVMTFNVNVKD-----ILNSR 1318
DB 1423 ENSYKRGMEELNFOCKEFDKRSRLYEYFRKMRDEIERKESQVLYLKETORKANLLLEAQ 1482
OY 1319 FNKRENRKNVL--ESDLIPK-----DLTSSNVVAVDPYFLNKKER-----RQKFL 1362
DB 1483 ANKLNIEKNITDEKELKAKKDVODIDSTNQKRELNLNBNKLLQOOSLIEREAI 1542
OY 1363 SSYNYIKDS-----IDTIDINPAND-----VLGYUYILSEKYKSD-----DSIKK 1402
DB 1343 NS-----KDSLNNKTIETIKROLHDKEMRVLRVLRBMKLAEBKYUTEINRLRTQTFDSEKQ 1598
OY 1403 -----YINDQGEN-----BKYLPF-----LNN 1420
DB 1599 DIKNFPPLFKINGNDMAFPYLPWLVPOQOQDNTLIQIRQLFOQOFMOQVRENELE 1658
OY 1421 IETLYKTIVNDKIDLVHLEKAKVLNITYEKSNNVAVKIKELNYLKTITQDKLADFKNNNFV 1460
DB 1659 LRQORNLLEKKLID--QIOLSEBOLNNKQSEPSKVSAMEKL--LEKTESRLNDFQKINY- 1713
OY 1481 GIADLSTDYNNHN 1493
DB 1714 -----LTKKVNQHN 1722

```
RESULT 45
T14867
Interaptin - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14867
R:Rivero, F.J.; Kuspa, A.; Brockamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142, 735-750, 1998
A:Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily in Dictyoc
ts.
A:Reference number: 218248; MUID:98365468; PMID:9700162
A:Accession: T14867
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1738 <RIV>
A:Cross-References: EMBL:AF057019; NID:g3549260; PID:g3549261; PIDN:AAC34582.1
C:Genetics:
A:Gene: abpd
A:Introns: 173/2; 1680/1

Query Match 5.2%; Score 438; DB 2; Length 1738;
Best Local Similarity 21.0%; Pred. No. 1.7e-06;
Matches 302; Conservative 254; Mismatches 528; Indels 352; Gaps 65;

OY 111 TNPSSSSDAKSYADLKHRRNRYLLTKELKYPOLFDTLHMLTLCNIGHFKYLIDG 170
DB 425 TNLQDQLTEQO-QOYQEKSLKVLNLELQE-KSNQLVKSNQSLTM----- 469
OY 171 YEINELLYKL-NFYDLRLAKLNDYCDAYCOIPFLKIRA-----NELDVLK 218
DB 470 QATNSLMKIGLMDLIDIPQDIKED--EIANLKIESKNLKCPODDNALQSR 527
OY 219 KLVEGYRKL-DNIKDNVGMEDYIKKNTIENINELIEESKRTIDKKNKATKEEKK 277
DB 528 SLTIEQTSOLDRKQILNELQERDQKLEFTSSNQLADNRVIDQLTN---EKQSI 583
OY 278 LVQAQVDSLTKQLE-EAHNLISVLEKRDPLKKNENIKELDKINEIKNPPASGNT 336
DB 584 TLOLOQDQDIKEFEKQQLLS---QIDSI--TTNIOEYODKFNMLQ---QEFNT 632
OY 337 PMLT-----LDKMKIEHEKEIKELAKTIKFNIDSLFTDPLELEYLR 380
DB 633 QQTINQOETHRLTQQLXQINTDYNKQLOSLQSIKD-NQITN-----EQLKQSL 661
OY 381 EKNKNIDISAKVETKESTENEPYNGVYPLSYND---INNALNE--LNS-FGDLIN 431
DB 682 EKDKETE---KLSNQEOQODEKINNMLLEIKKEDCLIERINQQLLENIDLNSKYQOLL 738
OY 432 PFDYTPKPSKNITDNERKKFIEKEK-----IKIEKK---IESDKSYEDR 477
DB 739 ETEENF-----LNSSEKEKQLELQSLQDERFNQINDELEKEKQSLIEDFENYKQ 793
OY 478 SKSLN-DITKEYEKL--LNEIYDSFNNNIDLTNFEKMMGKRYKVEKLTNNTFASY 533
DB 794 QLSNSNINQOQSLTIIEILSELKEQKELNDSKILIEKKOL-QOLOQEFQDLNKN---Q 848
OY 534 ENSKHNLEKLTALKTME-----DYSLRNIVVEKELKYKNT---ISKIENET 578
DB 849 KQHODQLLELEKQLOLOQEDYDQLENTNOSIENQNLNKNENNEKEQELQNL 908
OY 579 EFLVENIKKDEQLFEKKTITKDE--NKRPKEILEV-----SDIVKY 617
DB 909 NQOIEITQDPQOEFKSNISINIELVNKEKLTLOLOQDQDLKQONNSNDEKENDLIEK 968
OY 618 OVQKVLNKKIDELKKTQTLKKNVLEKHNHVPNSYKQENKQRPYLLIVLKEEIDKLKY 677
DB 969 ENQKLSIQNHELN---QLIEKN-ESDHK-----EQQLKQOSIENDLIEKENQ--- 1010
OY 678 MKPVESLINEEKKKNITQEQSDNSPSTEGEIT---GQATTKPQOQAGSALEGD--- 728
DB 1011 IOQLQSLNLEQRO-----QOSNQLSEKQDLNQLIEKNQFDQEQQLKQOSIENDLIEK 1064
```

```
OY 729 -----SVQAQAEQKQAPPPVPEAKQVTPPAPVNNKTEN----- 768
DB 1065 ENQIQLOSLNLEQROO-----SNQLSEKQDLNQLIEKNESDQEQQLKQOSI 1114
OY 769 -----VSKLDLEKLYEFLNTSYICHRYIIVSHSTNNEKILKOYKITEESKLS 820
DB 1115 ENLILKENQIQLOQLNLE---QROLOSEVSIIDDKILIELEKQLKQOSDILKLNDEK 1170
OY 821 -----DPLILEFNIONNI PVNYSFSDLSNLSQLEMEIYEKENVCNLKYL 866
DB 1171 QOQDQLODKQIREFDQLQTFNQFN--DKDSQFILOLQDQKQLOSLQDQ---LNQLKQ 1225
OY 867 KQNDKIKNLLEKAKKYSTSVKTLSSSSMQPLSLTPDK-PEVANDQSHS--TNLN--- 920
DB 1226 ENQEKQKQSEKDEK-----LOSIFQENQEKQKQSEKDEKQLOQONLNQLN 1273
OY 921 --NSLKFENILSLGKNKNITQELIGQKSENFYEKILKQSDPFYNESFTNPKRADDI 978
DB 1274 DENOEKVKQPSKDEKQSLQDQDLNQLKQENQEKQKQSEKQ-----EKQSLQOQL 1325
OY 979 NSLNDE--SKRKLLEDINKLKTQLQSLFDLYNKKYKILERLFDKKKRYGKYMQIKLT 1036
DB 1326 NQNLNDQIKKNEKILKE---KEEQDLKQODPNQOQSLQLEKLESEKENQLOQLKQ-- 1380
OY 1037 LKEQKLESKINSLNPK---HVLQNFVFPNKKRKEAIEAETNTLTKLLKHVGLVK 1093
DB 1381 -----ENETNQLNQOQSNELITQQLKQOLKQOQOQONNNEKETER-LIOEIQKQ 1433
OY 1094 YNGESSPLKTLSEESIQFEDNYSLENFVKLSLEKGLKDNLNLE--KKKLSYLSGLH 1151
DB 1434 QQELIDQSELN-KEIKIQT-----TQOEFQDLS--NHRSDQHLQLOQELDQLKQSP 1485
OY 1152 HLIAELKEVIKKNKYNYSNENNTDVNNALESY--KKFLPESTDVATVYSESGDTLEQ 1209
DB 1486 DQDHQFKKYL-DERYNLOLQLEQSTLSNNQDLQLEKELKPL-----ELDS 1530
OY 1210 SOKPKPASTHVGAESNNTTTSQNVDEVDVIVIPFGESEEDYDQGVATGEAVTPSV 1269
DB 1531 NEKQKITDILSNISNLQSLQNDKD-----LISEKNNSIKTLESITQQLSLIDE 1581
OY 1270 IDNILSKIEVEVLYLKLPLAGVYRSKLQLENNVMT-----FNVNVDKILNSRFNRE 1323
DB 1582 KDWLINDLQKQKQOQPPASSPSSPILSTPTPKPQRNQIIEIDLVEIYVRNQ 1641
OY 1324 NFKNVLESLLIPKDLTSSNRYVVKDPKFLNKEKRDKFLSSYV---IKSITQDI 1376
DB 1642 -----DLIR-----KNTKFKYKLENGDYIVNSIYRLSLDDNDSD 1678

RESULT 46
PQ0120
Major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (isolate B
N:Alternate names: gp195, PwMSA, PSA
C:Species: Plasmodium falciparum
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 09-Jun-2000
C:Accession: PQ0120
R:Kimura, E.; Mattea, D.; di Santi, S.M.; Scherf, A.
Gene 91, 57-62, 1990
A:Title: Genetic diversity in the major merozoite surface antigen of Plasmodium falci
A:Reference number: PQ0120; MUID:90382698; PMID:2205540
A:Accession: PQ0120
A:Molecule type: DNA
A:Residues: 1-144 <KIM>
A:Cross-References: GB:M32111; NID:g160490; PIDN:AAA29701.1; PID:g552211
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; surface antigen

Query Match 5.2%; Score 435.5; DB 2; Length 144;
Best Local Similarity 65.4%; Pred. No. 1.2e-07;
Matches 100; Conservative 8; Mismatches 24; Indels 21; Gaps 5;

OY 30 KRLALEDAVLTVGLSYLFQKRYVLENGTSGTAVTSTPGSKGVSAGSGSGS--VASGGSV 88
```


Db 2224 QLLATLSKIINKLN-----VKVSELDH-DEFSVNLISQLSKMNTEFNKKE----- 2269
Qy 1517 LDGLOGLMNLISOHCYKCCOQ 1539
Db 2270 ---QLSKILKTSIHLIINSDIPQ 2289

RESULT 49
T30822
Imp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infect Immun 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis P621 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID:95369882; PMID:7543881
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: EMBL:U01962; NID:q790243; PID:q790244; PIDN:AAA81013.1
C:Genetics:
A:Gene: Imp1
A:Genetic code: SGC3

Query Match 5.1%; Score 433.5; DB 2; Length 1365;
Best Local Similarity 21.1%; Pred. No. 1.8e-06;
Matches 318; Conservative 251; Mismatches 550; Indels 411; Gaps 74;

Qy 129 KHRV-----RNYLTTELKYPOLFDITNMTLTCDNHIGFKYLLDGYEBIN----- 175
Db 33 RHKJANOSOKENLLONEKKILO-KOLNELGYKIVNEIWFHEQVYLOSLKINKKSETKA 91
Qy 176 --ELLYKLNFPDLPRKLNQVANDCOQIPFNKIRANEVDYKLVFGYRKKPLDNKD 233
Db 92 IEEETLEKDAITLLISKIKN-----QINOKELEPAK-----FNEIKD 129
Qy 234 NVGKEDYIKK--NKKTIEINELIEES-----KKTIDKKNATKEEEKKLYOAO 282
Db 130 ---KLOGYIKNELSKOEYEHKONIEENELNKPISLESTLIEIQNATNNL----- 177
Qy 283 YDLSTYKOLEAHNLISVLEKRIDTLKKNENIELDKINELINPPRANGNPNTLLD 342
Db 178 --IKLNESTREKDN-----IDNLNAEQKLKASIQAOQL--POLSDN--DSEIA 222
Qy 343 KKKIEHEHEKIEIKIATKI--KENIDSLFTDPLEEYLRKKNKIDISAKYET-----KES 397
Db 223 KAKK--SLDAEIKNANOAVASNNTASMOASASSLD-----AKVAETIKKLETRKKDKA 274
Qy 398 TEPNEYPNGVYTPSYNDINNALNELNSF-----GDLINPFDTKEPSKNITYDN 447
Db 275 -----KPNELKOTRNQIOEFINTKNNPYSLLISQLT--SKDSKNSVYDS 319
Qy 448 ERKFFINEIKIEIKIEKKIESDKKSTEDSKS-----LNDITKEKILNIEIYSKFNNN 503
Db 320 SNKSDISANTELKQALAKANADYQADNLKSIKEQNLNSVSNANTLSAKLTJOK--DNT 377
Qy 504 ID--LTNFEKMMGKRYKYVEKLTHNHTFASYSKHNLE-KLTKALKYMEDYSLRNIVV 560
Db 378 IQAKTELEKVV-----QKADQAIKSNNTASMOSSKSSLDKAKVAEITKKLETFN-----K 427
Qy 561 EKELKYKNLISKIENEIETLVENIKKDEEQLFEKKITKDKNKDEKILLEVSDIVKVOVQ 620
Db 428 DKEKKF--NELKOTRNQIOEFINTKNNPN--YSELISQLTSKDSK----- 470
Qy 621 KVLAMKI--DELKQTQULILKVELKHNHVPNSYKQENKQEPYLLIVLKEIKDLKVMFP 679
Db 471 ---NSVTSDSNKSDIESANTEKQ-----ALAKANADRV----- 501
Qy 680 KVESLINEKKNIITEBOSDMSSEPTGEGITGATTKPG--QOAGSALEGSQAOAOQO 737
Db 737

Db 502 QADNLAKSIKEQL-----NNSVSNANTLSAKLTDKDNTIQOAKTELEKE-VOKADQAI 553
Qy 738 KOAQPPVVPVPEAKAQVPPAPVNNKTEKNSKLDYLEKIEYELNLSYICHKYLIVHS 797
Db 554 KSNV---TASQSAKSSIDAKVAEITKKLETFNK--DKAKNELKOTRNQIOEFINTKNN 609
Qy 798 TMN-EKILQYKITEESKLSQDPLDLLEFNQNNIPVMSMPSLNSLSQLEMEI-Y 855
Db 610 NPNVSELISQ--LTSKRDSKNSVTDSN-----KSDIESANTEKQALAKANA 655
Qy 856 EKEVACNLVK-LKD--NDKIKN-----LLEKAKVTSVKTLS 891
Db 656 DKVADNLAKSIKEQLNNSVSNANTLSAKLTDKDNTIQOAKTELEKELOKANQAIKSNNT 715
Qy 892 SSMOPL--SLTRPQ-----DKPEVANDTSHSTNLSNLKLEFNLSIGKKNKINYOELIG 944
Db 716 ASMOASKSSLAQVAEITKKLETFNKDEKAFNELKOTRNQIOEFINTKNNPYSSELS 775
Qy 945 OKSESENYEKLKSDTFYNESTFNPKSKADINDSLNDESKRRKLEEDINKLKTQLS 1004
Db 776 Q-----LTSKRDSKNSVTDSNKSNDIESANTEKQAL--- 807
Qy 1005 FDLNKKYKLERLFDKKKTVGKKMOJKITLKEOLESLNLSNNKHVLOPNSVFFN 1064
Db 808 ---NTAKA-----KSSID-----NELRPLKNDLSKRIEERGPIRNT--NFSWISS 848
Qy 1065 K-----KKEAIEAETE--NTLENTKILKHGYGLVYNGESSPLKTSSE--SIGTE 1113
Db 849 KLETTKNNLAELIKADIKNNPSSKQALDSSGOVOKLNL--LKTIEEFGKYVETK 906
Qy 1114 DNVALENFKY-----LSKLEGLKDNLNEKK-----LSYSSGLHH 1152
Db 907 NSNIGYRLFKLAQAEQFNNSDVKLKNAMEKQTLKSKOKLGNOSTKDYLTOLSTEST 966
Qy 1153 LIAELKEYIKKNVTGNSPENNVDVNNALSEYKFFLEGTGVATVYSESGSDTLEQSQP 1212
Db 967 QESTIKKIVNI-----QAHIRNNLSQYRLADKLI-----ANM 1001
Qy 1213 KKPASTHGAESNTITTSQNVDEVDVIVIFGESEEDYDGLGOVYTG-AVTPSYID 1271
Db 1002 KRQYGDVKIGES-----LOQMOLMDSDVL--SYDSTLKDDEPKALRYLVGQTKPPVSS 1055
Qy 1272 NLSK-----IEN-----EYEVLYKPLAGVYSKQLENNVMTFNVV----- 1311
Db 1056 WFINRNSIENYONLRLILVRENEILLDKA-----KDLKREKTIKEVDENINSIQD 1110
Qy 1312 -----KDLNSRKNRKNFNKVLIESDLIPYKDLTSSNVYVADPKFFLN-----KERR 1358
Db 1111 RAKRLKQELILNAK--NDLSNFTLNHQKNOFTAKDITPKISLENNKLNELINOYLLPIIKKA 1169
Qy 1359 DKPLSSYNYIKSDIDTDFINPANDVLYGYKILSEKYSKSDLSIKKYINDKQEGENEKYLPL 1418
Db 1170 VSKISEIEKKNEKELE-DIIRSENYFL-WKVEINKYISEL-----TKQVLSRNNIFE 1220
Qy 1419 NNIEITKYTV--NDKIDLFYHLEAKVLYNYTEKSNVEKIKELNLYKTIOQ--KLADKK 1475
Db 1221 NKWESIKQTDLNLKINENVSLLKEVIYN-----NSNAQYSINRI--LSTVPFIFIVAQOTR 1274
Qy 1476 NNNEVGADL 1485
Db 1275 SNNLRSLAEI 1284

RESULT 50
E71606
hypothetical protein PR0765v - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.;
Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.


```

Db 754 RLKSEI-----SHNENKMFSSKEGYAKIKELENNL-----787
Oy 800 NEKILKQYKIFKEESKSLSCDPLLEFIONNIPVMSMPDLSNLSQLFMEIYEREM 859
Db 788 -ERLSDLOSQIOLEIESTRSCDSOLK-AQNTI-----DDETKMKSLLEISNKT 838
Oy 860 VCNLYKLNDNDKIKMLLEAKKVSSTVKT-----SSSMQPLSLTPDCKEVSANDTSH 915
Db 839 TIE-----KLSEIENLDKRELKRTKFOYKFLDQNSDASTLEPLRLKEBIOQLD-----890
Oy 916 STNLNLSKLFPENILSLGNKNKIYQELIGQ--KSSENEYKTI-LKSDTFYVESFTYEVK 972
Db 891 ---ANSOIAQYEIIS--SNEHALIELKMLAKTKENYDAKIELEKKEKMAEEDLSRL 945
Oy 973 SKADINSINDESKR-----KLEBDIKKTKTQLOLSDLVKNYKYLKLERL-----1018
Db 946 GELGIRLQPLKGCALHPVOOSERLREVERIQMIE-----KIEKSTIVQ 994
Oy 1019 FDKKTVGKYYKQIKK---LTLLEKOLESKLSLNNPKHVLQNFVSFFNKKEAEIAETE 1075
Db 995 LCKKKNMGOYSTMKENNDLSFLVRLKADAC-----QAEILTKK 1036
Oy 1076 NLENTKTL-L-KHYGLVKKYVNGESSPLKTISESIQTEDNYASLENFVKSLKEGLKD 1134
Db 1037 SLSYSAODLLDKHER--KMEEKADYERELISNIEQTES--LRYENSVLKEIKVDTAAN 1091
Oy 1135 NLNLEKKLSTYSSGLHHLIAELK-----EYIKKNTGNSPSENNNDVNNALES 1184
Db 1092 NGDKHLVLVSFLSMRLRHRSNLETKLTCKRELAFVQKN--DSELTINDLQRTQTL 1148
Oy 1185 YKFFPEGTAVTVVSESQDLEOSOPKPASTHVGASNTI--TTSQNVDEVDVIT 1242
Db 1149 SEK---EVQCSAVIIDEFKDITKEVTO-----VNLKENNAILOSLSKNVTEKRE--I 1197
Oy 1243 VPICESEDDYDGLQ--VTGEAVTPSYIDNLSIKENEYEVLYLKLPLAGYRSLKQL 1300
Db 1198 YKOLNDROBEISRLQRLDIQTKEQV--SINSNKILYEESEMOCKOR---YQDLSQO 1250
Oy 1301 ENNVVTFVNVVKDI-----LMSRPNKRENFKNVLESDLIPKDLTSSVYVVKDPK 1351
Db 1231 K-----DAOKDIEKLTNEISDLGKLSAENANADLENK-----NRLKQAE 1295
Oy 1352 FLNKEKRDKF-----LSSYNTIKDSIDTDINFANDVLGYKILSEKYSDLISKYIND 1406
Db 1296 KLDASKQQAALTNELNELKAKDKLEODLHEN-----AKYIDDTKKAHELOSEVNS 1350
Oy 1407 KOGENEKTLPLNNIETL-----KYTVNDKIDLFVHLEAKVLYTEKSN-VEVKIKE 1459
Db 1351 RDHEKDTYRTIMEEIESLKKELQIFKTANSSSDAF---EKLKVMKEKDRIDERTKE 1406
Oy 1460 LNYLKTIDDKLADFPKKNPNFVGIALSTPDYNNNNLLTFLSTGVMFENLAKTVLSNLDG 1519
Db 1407 --FEKKLOETL-----NKSTSEAEYSKO-----IETLKKMELKEYEDE 1443
Oy 1520 NLQGLNLTISHOQYKQCPONSQCFRHLDERECKLLNTYKOE-GDKCVENPN--TCNEN 1577
Db 1444 TLRIRKEAENLKRIRLPSERIRQIKIISKRKE-----ELBEFPKRLKENAGSLTFIDN 1498
Oy 1578 NG-GCDADAKCTEEDSGSNCK 1598
Db 1499 KSGGEDAEELMNSPSKGNSSR 1520

```

RESULT 53

S41649 DNA polymerase - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 20-May-1994 #sequence_revision 26-May-1995 #text_change 09-Jun-2000

C:Accession: S41649

R:White, J.H.; Kilbey, B.J.; de Vries, E.; Goman, M.; Alano, P.; Cheesman, S.; McAleese, Nucleic Acids Res. 21, 3643-3646, 1993

A:Title: The gene encoding DNA polymerase alpha from Plasmodium falciparum.

A:Reference number: S41649; MUID:93376482; PMID:8367280

A:Accession: S41649
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1855 <WHI>
 A:Cross-References: EMBL:L18785
 C:Superfamily: Plasmodium falciparum DNA polymerase

Query Match 5.0%; Score 423; DB 2; Length 1855;
 Best Local Similarity 19.3%; Pred. No. 5.3e-06;
 Matches 385; Conservative 303; Mismatches 667; Indels 638; Gaps 88;

```

Oy 1 MKIIFLCSFLPF--IINTQCYTHESYQELVYKLEALBDAVLVGLSYLFOREKAVLNEGTS 58
Db 1 MYLIYLLKFPFFFSILNDV-----DKPEQVARNQGLD-----FLREKK-LKTLIS 47
Oy 59 GFAVYTTSTPQSKGSVASGSGSVASGGSVASGGSVASGGSVASGGSVRTNPDSNS 118
Db 48 GNMK-----SKEELHK-----TKNIEELMDQNTSS 72
Oy 119 DSDAKS-----YADLKH-----RVRYLLTIKELKYPQLEPLTNHMLTLC 158
Db 73 DSDGESRTKRRKRTRESYENSMYKMKSSCYLRSKDYENIYDMKTNEYN-----123
Oy 159 DNIHGFYLDGYEEIEINELLYKMFYFDLLRAKLANDVCANDYCOIPFNKIRANELDVLK 218
Db 124 -----KTIENNNNNN-----NYT-----YNNYNDMSFSNK-----DLEYSK 156
Oy 219 KLVFGYRKLPLNIKQNVGKMEYIKKKKKTIEININELIESKKTIDKNKNAFKEEKKLU 278
Db 157 -----YTK-VKNEEDIDKHQPY-----DIHDMGSSLKKKENYVQEOQSNM 198
Oy 279 YQAQVDSLTYNQLQJEANHLIS--VLEKRLDPLKKENIKELDKLNEIKNPAPASGT 336
Db 199 INVGNDDI-NEKKNHILAKNELVYKKEIQIKENEENIEKVEAGQSKNNNNKNNNDNN 257
Oy 337 PNTLLD-----KNNKIEBEKEIKE-----356
Db 258 NNNNVDDFYNDYILNEMNISPLFIKIENTETLKEEKKLKENDIKDEEDILENDISEML 317
Oy 357 -----IAKTIKFNIDSLFDDPLELEY 378
Db 318 KKKYEQELIKYEEBNNLFFVNVVYFDICKHNSIILFSGTLTKRKYSISIFTEINDRYYY 377
Oy 379 -LREKK-----NIDISA-KVETKESTEBN 401
Db 378 FLNKKKYVYENGEEIKFNEKFKIHMSEFLBEFKIIEYHNIKAKAKYIKRKNLNS 437
Oy 402 EYPNGVYPLSYNDINNALNELNSFGDLINPDYTERP-----SKN-----IYTONE 448
Db 438 SYDDELYIKVLVSYNNDRPIHEKFOKGSYLSFYCCMEDIVENFIKKNFKLPQWIKIKLU 497
Oy 449 RKKFLN-----ELKEKI-----KIEKKIESDKASYEDRSKSLNDITREYELKN 493
Db 498 RNDNLSNLTQCYEDCIVEDKKSILLDDCKVHEKRVKVTSPKSIDQTSPIKINNVSNNNNNN 557
Oy 494 EYDQSKFNNNIDLTPNEKMKMGKRSYKVEKLIHNNHFFAENSKNHL-----EKL 543
Db 558 NNNNNNNNNNNNNN-----NNNNNNNNNSNNGFYIISDQPKKI 600
Oy 544 TKALKYMEDYSLRNIVE-----KELKY-----567
Db 601 EGYQQLSDINLNLKYIKVVSLLNEENVHEIFSICSLVQIDKLKYINFGISRKSKKGT 660
Oy 568 -----KNLSIKIENE---IETLVENIKDEEOLF-----593
Db 661 MYNKKTKLCKLPDNEKELLHTELEKIKDIDIDYIGYNILNEDLEPLIRCHVHNLSYI 720
Oy 594 ---EKKITDEKKPKPEKILIEVDIVQVOYKVLNKKIDELKKTQILKLVNELKHNHVP 650
Db 721 LSRKKKKKKNE-----KKAKYK---FNGTNGSTGSMNIIIONIKGRILVYDI 763
Oy 651 NSYKQENKQEPYLLVYLKKEIDKLK-VFMPKVESLNEKKNI---KTEGQSDNSEPSTE 706

```

```

Db      764 LLCSNLSILFTY---CDEIIDIHVKNRQFODSKOYTNTKNTSTNKNTNKNTNYNQSVYSQQ   820
Oy      707 GEITGOATTKPEGQA-----GSALBGDSVOAOAEOBKQAOP-----VPPVPBPAKAQ-----   754
Db      821 SVVSQSOSVAV--SQOSVVSQSOSVAVTSOSTOSSQANOBNHNTOSTAHMHVATEEHIH   878
Oy      755 -----VTPPAPVN-----NKTENVASKLOYLELTYELMNTSYCHYLIVS--HS   797
Db      879 ASNRKATGHNDIPTDVASTNSHRSSWTHINIF-DELVANNILINCNSNI-HLYDAQSIITAN   936
Oy      798 TYNEXILKOYKITKEESKLSSCDPLDLFLFNIONNIPMYMSFSDSLNSLSQLFEIYEIK   857
Db      937 NLNFVYNSQIQCINE---TVNVCNLQII-----EKTRDILTKLSGYIMMR   978
Oy      858 EKVAC-----NLTKAKDNKIKNLL-----EAKKVST-----SVKTLSSSMOPL---   897
Db      979 SLICCTSERVEFFELLHEHYKKKFITPLIKRKTKLETOMKRNKNAKYLGGLVDPLCGY   1038
Oy      898 -----SLTPQDKPEVSANDTSHSTLNLSLKLFENISLIGKNKNLYOELOCQ   946
Db      1039 YDFELYLDNFNSLYSIITETWCFCSTJLKNCDVSTE--DKMLTKNGNI-----   1088
Oy      947 SSENFEYERILKSDPFYNESEFTNFKSRADDINSINDESRRKLEEDINKUKLTOLSFD   1006
Db      1089 --NNDYEKNI-HSDAEKNIHSDDDKNHSDCKANDNOXKYNKLED--NLLENWEI--D   1142
Oy      1007 LYNNKY-----LKLERLPDK---KTYGKYYKMQIKLTLKEOLESKLNSLNPKHVLQ   1057
Db      1143 FEDRSRPGILPCILKSLVEKRSVIKLLISNEKNKEKELLIQSLSIKLIS-NSIYGCLG   1201
Oy      1058 NFSVFENKKKEE--IAETEENTLETKTILKHVKOLVAYNGESSPLTLESBSIOTEDN   1115
Db      1202 MNNNRPVAKHLASYYTSGRNLOHTKFRKEEFNLKVIYDOTDISM-----IDTGIRK   1254
Oy      1116 YASLENFVYLSKLEBKLNDNLLEKKKLSYSSGHHHIAELEKEYIKKNKNTGNSPSENN   1175
Db      1255 ANNINYNESEFLAILINKINSINKNKKLEL--DLECIFSKLL-LLKKKRYACAKEYIDNN   1310
Oy      1176 TDVNNALESYK-----KFLPE-----GTDAVAVVSESBDTLEOSOP--KRDPAS   1217
Db      1311 -----LEKEYEMKGINFIKRDFSISKILGNEVLRITIF-TNRDVDSKNIPVPLENDS   1363
Oy      1218 THVGAESENTITTS-QNVDEVDDVLIIVPIFGESSEDYDD--LCGVYGEAVTPSVID--   1271
Db      1364 EOIHBYELPTINORINDDEFDDYYIITKRLKNHNEYODKNSLGHVLAERM--IKDGXY   1420
Oy      1272 NLSKIENEYEVLYLKPAGVYRSCLKOLENNVMFMVNDILMSREKNRENKENFVNES   1331
Db      1421 NICVAKKELOYCYCTSEDSARFKTKSEKLNSQCCFSINELIKKTYALKIDKEYIIRNOJLS   1480
Oy      1332 D---LIPYKDLTSSN-----VVVKDPYFLANKERDKFLSSVNYIKDSI-----DTD   1375
Db      1481 PINRCQOYIEGSAKSLSSCFNIYDKE-----IKTDKQEE-NYLETWVLSLINSND   1532
Oy      1376 INFANDVLCYKIIISEKTKSDSIDSTIKKIYINDOGENENEYLPPLNIEPLYTVNDKIDL   1435
Db      1533 BRFKDIHLKGFVACS---NCMHNVKPNIFIKFYFCGNCLTYLS----IQDIRNYIFSF   1583
Oy      1436 VTHLEAKVLANITYEKSANEVYKLELVNTLYQTOKDLADRKANNPNFADIADLVNHNHL-   1494
Db      1584 IHHLNTYTYKOLYICOGCTLKTRI-----FLKDKR-NCNPINCEYIKKNSLK   1629
Oy      1495 -----LTKFLSTGVNFENIAKTVLSNILDGNLOGM LN-----ISOHQCV   1533
Db      1630 PLISKKYIYILLLEYLF--LKDNL-KKIPSNLVEKKNSEQLNETGYODTNENDITSNKE   1686
Oy      1534 KKQCPONGSCFRHLDEREBCKLLN---YKQEGDKVENPWPCTENNNGGDADACATTEE   1590
Db      1687 EAKDENGSISINHDIINNEHSEKNHNHNYKK-----NNSNNNNNNNNKDIENEYCND-   1739
Oy      1591 DSGSGKKKITGEC   1603
Db      1740 -----FIYVLC   1745

```

Db	1071	SDL	-----	-	BSKNENPFKEKKAALNOLAKNESLILAK	-	KEKIDNKKELATTSKORD	11200
OY	483	DITFEYKLLNLEYDS	FENNIDITLFEKMMGRSVYKELTHNHTFASYENSKHLEK	542				
Db	1121	DAVSEHKITAELKETR	-----	IOLT	-----	EYKSNYOKIKEEYSNFORRETKOPOBK	1168	
OY	543	LTKALKMEDYSLRNIT	-	VERKELKYYKNLISKIE	-	NELFETLVENIKDEDELPEFKITK	599	
Db	1169	RNSLVEESLNSKIKELEARLSOBI	SILNOYLNKRISGVSYETNISTSTRS	-----	TS	1219		
OY	600	DENRPEKILTEVSDIYVEVQV	KVLLMNKKIDELKTKTOLIKNELKHNHIVPSKYOEKNQ	659				
Db	1220	YSDOP	-----	LDKEDILK	-----	KYDOLAFETIRN	-----	1247
OY	660	EPLYLYLKKEDIDKLV	FMKPVESLINEEKNKITEGOSDNSEPSGEBITGOATTKFGQ	719				
Db	1248	-----	LENEI	-----	EKKKNLISRLR	-----	FTETRL	-----
OY	720	OAGSALGDSGDOAOAOROKO	-----	AOPVAVPVEKKAQVPRPA	PVNNKTEWNSKL	-----	772	
Db	1270	-ASSFEDOKIKRAOMK	KLKLIDMDPSITPL	-----	DSILNEPLDNC	-----	DKESDIDKLMLE	1333
OY	773	-DYLEKLYEF	-LNTSYICHKYLIVSHSTNEKILKQYKITEEESKLS	SCDPLDLFNI	829			
Db	1324	VDYLKROLDIETRAHYDAE	MAISALHS	-----	KFRKI	-	OGESSLASSDIYKILKFEAS	1374
OY	830	-----	QNITPVMSMFDLSLN	-----	SLSQLFMEIYEKEMCNLYKLDN	869		
Db	1375	EERYKSLIEDKLT	KYPLDRKTNLVP	-----	GDIKNRDSISK	-----	YEEBI	RYKLI
OY	870	DKIKNLEAK	-KVSTSVKTLSSSSMOPLSTAPQ	-	DKPEVANSDDTSHTNLNSLK	-----	924	
Db	1423	YKLOEILNESNGKLSQ	TTDLRQSKSKALSSQDLRLOKDESPREQKELLSSTIKQOK	1482				
OY	925	-LFENILSLGKNNKNIY	DELIGOKSSSENFYEKLIKSDITYNESFTNFKYKSKRADDINS	983				
Db	1483	QOEFNCM	-----	DDLOGNLRUREH	HALKQAE	-----	EDVKNMASTIEKLT	1525
OY	984	ESKERRK	-----	LEEDINKLKTLOLSP	LYNNKYKLTLELFDKKRYTGAKYKMOAK	-----	1033	
Db	1526	ONQKKEKLINE	REMRNDSMOQETL	-----	TELKRYDVAKKIISDOLAHLKERTLSAVE	1580		
OY	1034	-----	KLTLKEOLESKINS	LNPKHVLONEVFEFNKKKAETAEENTLENTKILK	1086			
Db	1581	DRSOYTDEINRKLKEEL	NSLKAETNLK	-----	KEFAT	-	LKYKLTSTNDEAKISDILKOLD	1636
OY	1087	HYKLVAYYNGESSPLKT	SEESIOFEDMYASL	-----	ENFK	-----	VLSKLEKTLKDNLT	1138
Db	1637	HYTKVVEMLNNEKDAIS	LAEKELYQ	-----	KREALNECESLEKGR	IVSLTRKIOELESIDNO	1693	
OY	1139	EKKRLSYASSGLHHL	IAELKEVINKNNNYGNSSPENNNTDVNNALSES	KFLPREGTDVATV	1198			
Db	1694	KTDALQISNAL	-----	SSSTQKNKRTIKI	-----	KYLEE	-----	1724
OY	1199	VSEGSOTLEQSQPK	PASTHVAESNTTTTSON	-	VDEVDVYIIVPIFGSEEDYDGLG	1237		
Db	1725	-----	TL	-----	QLOMOUNSRNBEILVKTLOASNGYKDFD	-----	EKOKNIDILE	1766
OY	1258	QVNTGEAVTPSV	IDNITLSKIE	NEFELYLYLKPLAGVRSRLKQLENNWMTFNNVAVDILNS	1317			
Db	1767	E	-----	NQTLQKLTNDLOL	-----	QLKNLHERLSOTTEKNAML	-----	S
OY	1318	RFNRKREPNKVL	ESDLDIPYKDLTSSNYYVADPKYPLNKKERDKPLSSYNYIKDS	IDTD	-I	1376		
Db	1801	KIHLELNNVVS	-LETDL	-----	KYEEBKKNKSLERAVEELQTKNSQOQTVI	1844		
OY	1377	NFADVNGYYKII	SEKKSDSDSKIKKYINDKOGENSEKYLPLNLTNLETLYKTYNDKIDFLV	1436				
Db	1845	ELAKKNSRFEDEALKEA	DISPLEKYYISOOELEMKKSTRONSYSYRDKVQAEKDAETEFWK	1904				
OY	1437	IHLAEKVL	-NYTEKSNVENEKI	1457				

DB 1905 SRYESTMIGSKNIDSNNAQSKI 1926

RESULT 55

T18440

hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 09-Jun-2000

C:Accession: T18440

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18935

A:Accession: T18440

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-4550 <LAMB>

A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB11121.

C:Genetics:

A:Map position: 3

A:Note: C0425w

Query Match	5.0%;	Score 420.5;	DB 2;	Length 4550;
Best Local Similarity	19.1%;	Pred. No. 1.8e-05;		
Matches 389;	Conservative 288;	Mismatches 679;	Indels 685;	Gaps 89;

Oy	72	SVAGSGGGSVAASGGSVAASGGSVAASGSGCSGRRTNSDMSSSDPAKSYADLKHR	131
Db	1693	SCSSRSFSGSLYNTSDNISESEISNNLSPSSPSKSEN-YCSSISDHEIKEDVTYIRNTHY	1741
Oy	132	VRNLTITKELKYPOTEDLN-----HMLTLCDN-----	162
Db	1742	IINBMKDGKKNYSGSKKADMSNKKGGDKSCAKNNICOMKKNIYDNKKLIYDNKSSDPADAVE	1801
Oy	163	GFKLIDGEEINELLYKLNFFYDLRAKLDANCANDYQIIPFNKI-----RANELDVL	217
Db	1802	NIDHIE-KEEIDILIMNLGLYNYLNKNNNN---NYVMK---TITKLCTGHSNEHDT-	1853
Oy	218	KLVEGYKRPDLNIDKDVGMEDYIKNNKKTENI-----ELI	256
Db	1854	-----KAVMNKILNT-----LNKKYNNINCIIFYHYNSFLSLFSDSEYIELI	1897
Oy	257	--EESKTTIDKN-----	267
Db	1898	RKEERKEDIKEKIKRGKYCHOEDRDDNNNNNDNNNDNNNDNNNDNNNDNNNDNNND	1957
Oy	268	-----NATKEEKKK-----LYOAOYDLSYKOLEZAHNLISYLEKRIPTLK	310
Db	1958	NNNDNNSGEDOLINEBNNKRRKINNLSTIEHHHNSDHHKEPKKKKKSNETQINDDTNK	2017
Oy	311	KNENIKEL-----LDKINEIKNPPRANGTPTNLT-----DNKKIEEHEKEIK	355
Db	2018	KETHNEIDHKEOKOTCDNGIEKKDOQYSKNIIISKYHFEEFKNNIKKDEHKEYN	2077
Oy	356	ELATITFNDSLTPOLELEYLRKKNKNIDISAKVEKESTEBENEPNCGVTPPLSND	415
Db	2078	ERNDKMS-ECMMLONKRDNDNNNNNNNNDDNNNNNNNDNNNDNNNDNNNDNNNDNNNN	2136
Oy	416	INNALNELNSFGDLINFDYTKEPSPKNITYDNERKKFLINELKEK-IKIEKKIESDKSY	474
Db	2137	NDDNNNNNN---NDDNNKRPD---LDGFTSYIDEINQKEINIFSKYIDKIIENV	2187
Oy	475	EDRKSILNDITKEYEKILNFIYD---SKFNNNIDLTFEKKMGKGRYSKYVEKELTHNTF	530
Db	2188	ETKCIKMNVCDDGIYILIRENGSLTCQKVNNNI---IKNMQEKERYKKRKKONKNN--	2241
Oy	531	ASYSNESHNLNEKLTALKAYMEDYSL-RNIVYERELKYKKNLISKIENIEFTLVENIKDE	589
Db	2242	---HNKKMKN-----KMYIDININNNVPDIQKRN---INNNIININNNN	2281
Oy	590	EQLFEKKIYDENKPEDEKILIEVSDI-VKQVOQVLLM-----NKIDELAKT	634
Db	2282	NNNSNSHSHNNNNNSHN--INISDLKKKININIKYUULFYRLPLFEGDKYINDIDVAKSI	2339

Best Local Similarity 20.0%; Pred. No. 1.1e-05;
Matches 347; Conservative 283; Mismatches 588; Indels 516; Gaps 87;

```
OY 2 KIIFFLCSTLEFIITMOCVTHESYOELVAKLEDAVLGT--SIFQEKVNLNGTSG 59
Db 635 KIIFF--CELYIYF-----KILKRIOSLDMMKINHEVKGMDKVLN----- 667
OY 60 TAVTSTPGSKSVASGSGSVASGSGSVASGSGSVASGSGSGNRRTNPSNSD 119
Db 668 -----NSYKCYDKNDYGSFSPYNTKS 689
OY 120 SDAKSYADLKHRVRYLLTIKELKYPQLDNLHMLTJLDNIHG-----EKYLDGYE 172
Db 690 NDYKSH-----YIKKKMKNVSVQCSNESIILKEQENEEKKKKKKKMMENTFL 738
OY 173 EYNELLYKLNFEYDILLAKLNDVANCANDYQIIPNLIKIRANEDVLKLVG----YRKL 228
Db 739 NNNMLTYNNINFEFDLI--INERGNFQFPYNNIKKKRQKNEKGLEMMYNNIFOLYMKYI 795
OY 229 DNKDNVGMEDYIKKKKNTIEN-----INELIESKKTIDKNNKATKEEK 275
Db 796 LNEPSKFEFLKFLKNVNDINTNSTIYNKYIIN--MYYHRKDCFEKQIHSKHHM 854
OY 276 KLYOAYDLSYKQLEBAHNLISYLEKRIDTLKKKENIKELDKINEIK--NPPANG 334
Db 855 KRIH-----LRDKFIEYEKEN-----ELIDCNMNMNKKKEIN 889
OY 335 NPPNLLDKNNKIEHEKEIKELAKTIKFNIDSFTDPLELEYLYREKKNDIS--AKYE 393
Db 890 NNYNNKIDNNNIEIDMSNNFIETYYIIFLLNMDTYIOFLYYL--KNTYILFSVAKA 947
OY 394 TKES-----TEPNYPNGVTPPLSYN--DINNALNELNSFGD--LNPFDYTEPSKNIT 445
Db 948 EBNLSMLKTLKRNKHYIKLRNHHIINSVYKILN--NYKKDEIFLYDITK-----WT 999
OY 446 D---NERKFEIIEKIKIEKK-----IESDKSYDRS-----KSLDITKEYE 489
Db 1000 EKMCMTTDLIYNDVKNFTIDLENDIDPIITNDKEEYHANNISILVKKHNSVYKLLK 1059
OY 490 KILNEIY--DSKFN-----NNIDLTNEKMMGKRYSVKELTTHNHFASVENSKNL 540
Db 1060 KIKNSIILKDLKLCNFTINKYIHTNTYDK--HNKIYDOKIKMWTY---PNNKKKVV 1114
OY 541 EKLTRALKYMEDYSLR--NIVEKELKYK-----NLISKIENETIETVE--NIKKD 588
Db 1115 KIKKFIISYDAYIYHGVALNLFNNAAYEKLSONPPSSIDLKKEGQNNYIINGEIKKY 1174
OY 589 EEO-----LEKKITKDEKPRDEKILEVSDIVKVOYQVLLMMKIDELKKTOLI 637
Db 1175 EEONNFIIRPNINISGKMLSCHNKTNSNTLOGND-----REANILDADEBRLKRNKI 1230
OY 638 LKNVELKHHIIV--PNSYOENQOEPIYLLVAKKELDKLVPRKVESLINEEK--KNIK 693
Db 1231 ONRAVYQOMTJNNNSKKRRNMQTEYY---KKKEKKRYDQKNOQ--TNEQCAQKNOQ 1284
OY 694 TEGQSDNSEPTEGETGATTKPGQAGSALGSDVQAQAOBKOAOPRPVPEAKA 753
Db 1285 TNEQ-----KNDQTNQTDQTDQ----- 1304
OY 754 OVPTRPAPVNNKTENVSKLDYLEKLEFLNTSY--IC-----HKYI-----LVS 795
Db 1305 -----TNDQTDQTN--DKIKRFYKNITCYCYKLMCKREYSNKKYLSWLCSGMSLDVYI 1355
OY 796 HSTNMEKILKQYKITEESKSLSCDPLDLFNIONNIPVMTSMFDSLNNIS--QLPMEI 854
Db 1356 NFIIIVRLYEKY--NKNENKTKC-----PIPRILLYLTHOSSILTSFSCVGI 1400
OY 855 YEKEM-----VCNLYKLDNDKIKNLEBAKKVTSVTLSSSQPISLTPODPR 905
Db 1401 RKDKDKIIPFPASFISLELHIK--KKIKIKLSNKLNVSNNSKSYCSNKNYNI--MKEEKK 1458
OY 906 EVSANDDTSHSTNL-----NNSLKLFEENILSIGKKNKIYQELIGKSSENEYETIL 956
Db 956 -----NNSLKLFEENILSIGKKNKIYQELIGKSSENEYETIL 956
```

```
Db 1459 HASSRSVHVNQTDRTDVLSEIFYHNNTANIF-----CCKDCVWK--VRETENEKKEKOR 1511
OY 957 KQSDFFYNSFPNFVSKADINDSLNDESKRKLEEDINKIKKTLQSLDLYN----- 1009
Db 1512 KMK--KPMENENENAVIK--DDEKNTYNIILK--NINENIDK--KKSININICINDIPTNVA 1565
OY 1010 --KYLLELELEDKRRKTVGKYKQWIKKTLLEQOLESKLNSLNPNPHVLQNFVFPNK-- 1065
Db 1566 NKKEYSLPRKCLNK-----INDFK--NLFYLLCYKN-----NNIDLIQLYDCLNNNY 1612
OY 1066 --KKEAIEAETENTLENTKILIKH-----YGLVYYVNGESSPLTKLSESIQTEDNYA 1117
Db 1613 THIKKNMQKLEKG-----KHGRNFGYGFVFTPNNSVPLTKKRLKLLKKNY-- 1659
OY 1118 SLENKVLKLBCKLKDNLNLEKKKLSYSSGJHLHIAELKEVIRKKNKNTGNSPSPNND 1177
Db 1660 -----KGNKKDKEDDN--NYHNDKNNY--SDNLEFYNHDTNN--NNNNNNNNNNNNNN 1709
OY 1178 VNNALSEYKRFLEPGTDVATVVSSESGSDPTLEQOPKRPASTHVGAESNTITTSQNVDEY 1237
Db 1710 NNICLKNKNNNTI-----MHEDINANKRESLAKKKKK-----KKNCIQKNNNICERK 1756
OY 1238 DDYI-----IYIFGESEEDYDLAGOVVTEGAVTPSVIDNLSKIENEYEVLYLK 1287
Db 1757 KSNIHNNSSKYIFNTVRFF-----KMKDIKAKINTKCKD---ENSISCIINNRE----- 1802
OY 1288 PLAGVYRSILKOLENVMTFNVNVKD--ILNSRPNKR-----ENFKVNLSEDLIPYKDLTS 1341
Db 1803 -----KRNIFKLNRLNIFLNNNSNDKYANAYIYNSNVTNYGKRYKILNKDV-----H 1850
OY 1342 SNVYVDPKPYFLNKRDRKFLSSYNYIKDSIDTIDIFANDVLGYKYLSEKYSKDSLSIK 1401
Db 1851 INNILLHTYK--QKKKKSTYIIS-----DNNNNNNNAEDDISRKL---KFRDIGKNTK 1901
OY 1402 -KYINDKQGENEYUPLNLIETLYKYVNDKI---DLFYIHEAKVLANITYEKSNEVKI 1457
Db 1902 QXYINDH-----NNINSYDNNINNGLLNEHKRYVLHNECK----- 1935
OY 1458 KELNYLKTQDKLADFKKNNNFVGIADLSTDYVHNHMLTKFLSTGWFENLAKTVLSNLL 1517
Db 1936 -----NKNQITG--YSIKYKN-----VSEKSCSDVITSLK 1966
OY 1518 DGNLQGLMNSHQCVKQKOPNSGCFRLHDEREC--KCLLNYKQSGDKVEMP 1570
Db 1967 DKIKRKR-----KKKIQKKN---YENENIVCLDCLISYKMKMLRITYGNP 2007
```

RESULT 58

A:56539

N:Alternate names: macrogolin

C:Species: Homo sapiens (man)

C:Date: 19-Oct-1995 #sequence-revision 26-Jan-1996 #text-change 10-Dec-1999

C:Accession: A56539; 637536

Mol. Cell. Biol. 14, 2564-2576, 1994

A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein

A:Reference number: A56539; PMID:94187728; PMID:7511208

A:Accession: A56539

A:Molecule type: mRNA

A:Residues: 1-3259 <SEE>

A:Cross-references: EMBL:W75304; NID:g405714; PIDN:CA53052.1; PID:g405715

C:Genetics:

A:Gene: GDB:GOLGB1; GCP: GCP371

A:Cross-references: GDB:454958

A:Map position: 3q13.31-3q13.31

C:Superfamily: giantin

C:Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TM>

Query Match

Best Local Similarity 4.9%; Score 412; DB 1; Length 3259;

Matches 314; Conservative 301; Mismatches 592; Indels 376; Gaps 72;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 07:21:31 ; Search time 26 Seconds
(without alignments)
1854.776 Million cell updates/sec

Title: US-09-269-874a-3
Perfect score: 8424
Sequence: 1 MKRIFFCISFLFIIMQCV.....SNFGISFLILMLIVSFI 1639

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 180 summaries

Database :
1: /cgn2_patents_AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	%	Length	ID	Description
1	1824.5	21.7	377	4	US-08-195-705-5	Sequence 5, Appl1
2	1763.5	20.9	375	4	US-08-195-705-3	Sequence 3, Appl1
3	1289.5	15.3	394	4	US-08-195-705-2	Sequence 2, Appl1
4	1254.5	14.9	394	4	US-08-195-705-4	Sequence 4, Appl1
5	615	7.3	115	1	US-08-152-922A-7	Sequence 7, Appl1
6	448	5.3	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
7	387	4.6	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
8	356.5	4.2	3248	5	US-08-353-700-1	Sequence 1, Appl1
9	356.5	4.2	3248	5	PCT-US95-16216-1	Sequence 1, Appl1
10	353.5	4.2	2285	4	US-09-308-375-2	Sequence 2, Appl1
11	351	4.2	1786	4	US-08-973-462-8	Sequence 8, Appl1
12	350.5	4.2	2482	4	US-08-328-354-6	Sequence 6, Appl1
13	343	4.1	1939	4	US-09-310-187A-1	Sequence 1, Appl1
14	338	4.0	1886	4	US-08-938-105-3	Sequence 3, Appl1
15	332	3.9	976	4	US-09-104-324B-4	Sequence 4, Appl1
16	314.5	3.7	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
17	310	3.7	1211	4	US-09-134-001C-4820	Sequence 4820, Ap
18	308	3.7	1312	2	US-08-687-080-51	Sequence 51, Appl
19	307	3.6	1312	2	US-08-592-126-148	Sequence 148, App
20	303.5	3.6	1038	4	US-09-541-782-4	Sequence 4, Appl1
21	303.5	3.6	1038	4	US-09-723-820-4	Sequence 4, Appl1
22	302.5	3.6	3169	4	US-09-453-702B-257	Sequence 257, App
23	301	3.6	53	1	US-08-290-919-3	Sequence 3, Appl1
24	298.5	3.5	2710	1	US-08-480-604A-6	Sequence 6, Appl1
25	298.5	3.5	2710	2	US-08-405-496A-6	Sequence 6, Appl1
26	298.5	3.5	2710	2	US-08-915-136-6	Sequence 6, Appl1
27	298.5	3.5	2710	4	US-08-957-310-6	Sequence 6, Appl1

28	298	3.5	3135	1	US-08-323-170B-2	Sequence 2, Appl1
29	298	3.5	3135	4	US-08-954-441-2	Sequence 2, Appl1
30	296	3.5	1588	5	PCT-US93-07261-11	Sequence 11, Appl
31	296	3.5	1663	5	PCT-US93-07261-16	Sequence 16, Appl
32	294	3.5	1388	4	US-09-572-191-2	Sequence 2, Appl1
33	294	3.5	1388	4	US-09-723-262-2	Sequence 2, Appl1
34	294	3.5	1388	4	US-09-723-219-2	Sequence 2, Appl1
35	293.5	3.5	1010	4	US-09-134-001C-5178	Sequence 5178, Ap
36	293.5	3.5	2391	2	US-08-446-855A-2	Sequence 2, Appl1
37	293.5	3.5	2391	4	US-09-150-741-2	Sequence 2, Appl1
38	292.5	3.5	3418	4	US-08-755-587-44	Sequence 44, Appl
39	288.5	3.4	3418	2	US-08-603-753D-4	Sequence 4, Appl1
40	288.5	3.4	3418	4	US-09-039-753-4	Sequence 4, Appl1
41	288.5	3.4	3418	4	US-08-986-106-4	Sequence 4, Appl1
42	282	3.3	53	1	US-08-290-919-4	Sequence 4, Appl1
43	280	3.3	48	1	US-08-290-919-12	Sequence 12, Appl
44	279.5	3.3	1589	3	US-09-356-952-4	Sequence 4, Appl1
45	279.5	3.3	2329	3	US-08-755-587-16	Sequence 16, Appl
46	279.5	3.3	3418	2	US-08-639-501-2	Sequence 2, Appl1
47	279.5	3.3	3418	3	US-09-044-946-2	Sequence 2, Appl1
48	279.5	3.3	3418	3	US-09-044-908-2	Sequence 2, Appl1
49	278.5	3.3	2366	1	US-08-480-604A-10	Sequence 10, Appl
50	278.5	3.3	2366	2	US-08-405-466A-10	Sequence 10, Appl
51	278.5	3.3	2366	4	US-08-915-136-10	Sequence 10, Appl
52	278.5	3.3	2366	4	US-08-957-310-10	Sequence 10, Appl
53	278.5	3.3	3788	4	US-09-336-447A-76	Sequence 76, Appl
54	275.5	3.3	1093	4	US-09-315-793-52	Sequence 52, Appl
55	274	3.3	48	1	US-08-290-919-1	Sequence 1, Appl1
56	274	3.3	1354	3	US-08-685-871-2	Sequence 2, Appl1
57	272	3.2	1535	3	US-08-755-587-185	Sequence 185, App
58	271	3.2	48	1	US-08-290-919-2	Sequence 2, Appl1
59	271	3.2	1164	4	US-08-923-992A-10	Sequence 10, Appl
60	271	3.2	2154	2	US-08-841-349-4	Sequence 4, Appl1
61	270	3.2	2052	2	US-08-790-912-2	Sequence 2, Appl1
62	269	3.2	1435	2	US-08-568-459A-4	Sequence 4, Appl1
63	269	3.2	1435	2	US-08-487-826B-4	Sequence 4, Appl1
64	269	3.2	1435	2	US-08-487-826B-4	Sequence 4, Appl1
65	267.5	3.2	1388	2	US-09-210-288-4	Sequence 4, Appl1
66	266.5	3.2	1579	3	US-08-755-576-4	Sequence 4, Appl1
67	266	3.2	1164	4	US-08-923-992A-2	Sequence 184, App
68	265	3.1	1363	1	US-08-425-061-23	Sequence 23, Appl
69	265	3.1	1363	2	US-08-825-886-23	Sequence 23, Appl
70	265	3.1	1852	1	US-08-425-061-24	Sequence 24, Appl
71	265	3.1	1852	2	US-08-825-886-24	Sequence 24, Appl
72	265	3.1	1863	1	US-08-425-061-16	Sequence 16, Appl
73	265	3.1	1863	1	US-08-480-784-2	Sequence 2, Appl1
74	265	3.1	1863	1	US-08-483-553-2	Sequence 2, Appl1
75	265	3.1	1863	1	US-08-487-002-2	Sequence 2, Appl1
76	265	3.1	1863	1	US-08-483-554B-2	Sequence 2, Appl1
77	265	3.1	1863	1	US-08-488-011B-2	Sequence 2, Appl1
78	265	3.1	1863	2	US-08-825-886-16	Sequence 16, Appl
79	265	3.1	1863	4	US-08-850-727-2	Sequence 2, Appl1
80	265	3.1	1863	5	PCT-US95-10202-2	Sequence 2, Appl1
81	265	3.1	1863	5	PCT-US95-10203-2	Sequence 2, Appl1
82	265	3.1	1863	1	PCT-US95-10220-2	Sequence 2, Appl1
83	265	3.1	1864	2	US-08-790-912-3	Sequence 3, Appl1
84	264	3.1	1863	1	US-08-798-691-4	Sequence 4, Appl1
85	264	3.1	1863	3	US-08-825-487A-4	Sequence 4, Appl1
86	264	3.1	1863	4	US-09-074-476-6	Sequence 6, Appl1
87	263	3.1	1863	1	US-08-598-591-2	Sequence 2, Appl1
88	263	3.1	1863	1	US-08-798-691-2	Sequence 2, Appl1
89	263	3.1	1863	1	US-08-798-691-6	Sequence 6, Appl1
90	263	3.1	1863	3	US-08-825-487A-2	Sequence 2, Appl1
91	263	3.1	1863	3	US-08-825-487A-6	Sequence 6, Appl1
92	263	3.1	1863	4	US-09-074-476-2	Sequence 2, Appl1
93	263	3.1	1863	4	US-09-074-476-4	Sequence 4, Appl1
94	262.5	3.1	106	1	US-08-290-919-11	Sequence 11, Appl
95	260	3.1	740	1	US-08-257-073-5	Sequence 5, Appl1
96	259.5	3.1	1388	2	US-08-685-576-1	Sequence 1, Appl1
97	257	3.0	1332	4	US-09-345-882-29	Sequence 29, Appl
98	254	3.0	1863	2	US-08-603-753D-2	Sequence 2, Appl1
99	254	3.0	1863	4	US-09-099-753-2	Sequence 2, Appl1
100	254	3.0	1863	4	US-08-986-106-2	Sequence 2, Appl1

101	284	3	0	1863	4	US-09-007-67883-49
102	253	3	0	3289	2	US-09-477-851-2
103	251.5	3	0	1169	4	US-09-265-829-20
104	251	3	0	793	6	5231168-2
105	250	3	0	1128	4	US-08-933-992A-6
106	249.5	3	0	1494	3	US-08-755-587-186
107	249	3	0	1147	3	US-08-470-260-5
108	249	3	0	1147	3	US-08-471-491-5
109	249	3	0	1147	4	US-08-466-662-5
110	245	2	9	1161	1	US-08-053-614-4
111	245	2	9	1181	1	US-08-316-978B-4
112	245	2	9	1181	2	US-09-034-306-4
113	245	2	9	1181	4	US-09-269-437-4
114	245	2	9	1181	5	PCT-US93-08782-4
115	244.5	2	9	1104	4	US-08-923-0928A-4
116	243.5	2	9	4556	4	US-09-180-422B-27
117	242.5	2	9	652	4	US-08-559-868B-2
118	242	2	9	1664	1	US-09-599-652-2
119	242.5	2	9	1664	2	US-08-642-846-2
120	242.5	2	9	1664	4	US-09-264-604-2
121	241	2	9	872	1	US-08-766-014-2
122	241	2	9	1683	3	US-08-755-587-183
123	239.5	2	8	1073	4	US-09-541-782B-6
124	239.5	2	8	1073	4	US-09-723-820-6
125	239	2	8	1088	4	US-08-923-922A-8
126	238.5	2	8	872	3	US-08-851-843A-8
127	238.5	2	8	872	3	US-08-851-843A-54
128	238.5	2	8	872	4	US-08-974-549A-221
129	238.5	2	8	872	4	US-08-854-050-8
130	238.5	2	8	872	4	US-08-854-050-54
131	238.5	2	8	872	4	US-09-430-323-8
132	238.5	2	8	872	4	US-09-430-323-54
133	238	2	8	2101	1	US-08-135-487-4
134	238	2	8	2101	1	PCT-US93-06160-4
135	238	2	8	2101	1	US-08-466-390-4
136	237	2	8	2101	1	US-08-470-950-4
137	237	2	8	2101	1	US-08-467-781-4
138	237	2	8	2101	2	US-08-483-924-1
139	237	2	8	2101	4	US-09-452-294-4
140	236	2	8	1861	2	US-08-790-912-4
141	235	2	8	1202	1	US-08-435-061-22
142	235	2	8	1202	2	US-08-825-886-2
143	234	2	8	1118	4	US-09-379-523-3
144	233.5	2	8	1116	4	US-08-914-929-6
145	233	2	8	983	3	US-09-134-001C-3814
146	233	2	8	1589	3	US-08-735-587-189
147	232	2	8	912	4	US-09-134-001C-2993
148	230.5	2	7	1338	1	US-08-471-033-50
149	230.5	2	7	1338	2	US-08-471-044-50
150	230.5	2	7	1338	2	US-08-463-483A-50
151	230.5	2	7	1338	2	US-08-471-046A-50
152	230.5	2	7	1338	2	US-08-470-568B-50
153	230.5	2	7	1338	2	US-08-469-324-50
154	230.5	2	7	1338	3	US-09-300-529-50
155	229.5	2	7	2314	4	US-09-268-347-49
156	228	2	7	878	4	US-09-134-001C-4378
157	228	2	7	885	2	US-08-533-306A-4
158	228	2	7	885	2	US-08-742-923A-4
159	228	2	7	2938	5	PCT-US94-00198-3
160	227.5	2	7	840	4	US-08-974-549A-190
161	227.5	2	7	1315	4	US-08

[illegible]

	174	219	2.6	1346	2	US-08-471-044-23	Sequence 23, Appl
	175	219	2.6	1346	2	US-08-463-483A-23	Sequence 23, Appl
	176	219	2.6	1346	2	US-08-471-046A-23	Sequence 23, Appl
	177	219	2.6	1346	2	US-08-470-566B-23	Sequence 23, Appl
	178	219	2.6	1346	2	US-08-469-334-23	Sequence 23, Appl
	179	219	2.6	1346	3	US-09-300-529-23	Sequence 23, Appl
	180	218.5	2.6	631	4	US-08-477-831C-11	Sequence 11, Appl

ALIGNMENTS

```

RESULT 1
US-08-195-705-5
: Sequence 5, Application US/08195705
: Patent No. 6420523
: GENERAL INFORMATION:
: APPLICANT: Chang, Sandra
: APPLICANT: Hul, George
: APPLICANT: Barr, Philip
: APPLICANT: Gibson, Helen
: TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM
: TITLE OF INVENTION: FALCIPARUM VACCINE
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Davis Hoxie Faithfull Hapgood
: STREET: 45 Rockefeller Pl.
: CITY: New York
: STATE: N.Y.
: COUNTRY: USA
: ZIP: 10111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/195,705
: FILING DATE: 14-FEB-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Jacobs, Seth H
: REGISTRATION NUMBER: 32140
: REFERENCE/DOCKET NUMBER: 11880A3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-757-2200
: TELEFAX: 212-586-1461
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 377 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: NO
: FRAGMENT TYPE: C-terminal
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium falciparum
: STRAIN: WEL
: US-08-195-705-5
Query Match 21.7%; Score 1824.5; DB 4; Length 377;
Best Local Similarity 93.6%; Pred. No. 8.8e-90;
Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;
```

QY	1264	AATPSVINDILSKIEENEVEYLKPLAGYYRSLSKQLENNVTFNVNKKDIINSFPNKR	1323
DB	1	AATVSVIDIILSKIENEXEVLTKPLAGYRSLKQLENNVTFNVNKKDIIINSFPNKR	60
QY	1324	NKNVLESLLPYKDLTSSNYVKDPYFLNKKEKDKFLSSINYIKDSIDTDINRANVYL	1383
DB	61	NKNVLESLLPYKDLTSSNYVKDPYFLNKKEKDKFLSSINYIKDSIDTDINRANVYL	120


```

Query Match      5.3%; Score 448; DB 4; Length 10182;
Best Local Similarity 20.6%; Pred. No. 6.9e-15;
Matches 364; Conservative 250; Mismatches 651; Indels 498; Gaps 81;

QY 107 NSRRNPSSNSDSDAKSLADLKHVRYNLTITREKLYPQLF-DLTNMLTLCDN---- 160
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7957 NKAKALNDMMKLIKIDIVAQDQNVQSNRY---INEDSPQNNYNDTIINHAQSIIDQVANP 8013
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 161 -----LHGFTYLI---DGEYELNELLKLFYEDLLRAKLANDVCANDYCOIPEML 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8014 TMSHDEIENALINIKHAINALDGEHKLQAKENAN---LTINSND----- 8056
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 KRAEMLDVLAKLV--GYRRKPLDNIKQNVGCMEDYIKNNKTTIENINELLEESKKTIDK 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db	8057	-LNAFOAIRLRYNEQOTREKAEAOLOSQAOLNDAMKHLRNSIONQSSVROESKTY----	8112
Qy	266	MKNATREEEKKRLYOAOYDLSIVYKNOLEEAHNLISYLEKRIDTLKKNENIKELDKINEI	325
Db	8113	NASDAQKFO-----YHAAVEVENTIINBQHPTLD-----KEIIKOLUTGVNO-	8154
Qy	326	KNPFPANGPTPTLLDKKKIEE-----HEKEIKELAKTYIKRIDSLEPDL-	373
Db	8155	-----ANNDLGVELLDADQONAHOSIPTLMLHNOQOALNEKINNATVREVAIIQO	8209
Qy	374	-----ELEYLKRNKNIDIASVETREESTEPENRNCVTY-----PLSY	413
Db	8210	AKLLDHAMENLEESIKOK-EQYQOSSYVINBESDVQEPYDADVAHYTELINOTVPTLSI	8268
Qy	414	NDINNALNELNSF-----GDLINPDYU-KEPS-NITYONERKFFINEI-KEKIKIE	463
Db	8269	EDIEHAIEVQAOKKOLRGKOKUYOTIDLADKELSKLDLTQSOSISSINOJTYAKTTE	8328
Qy	464	KKKIESDKSYEDBSKSINDITREKLELEIYDSKFNINNDITNEPKMKGRYSYKEX	523
Db	8329	VAQAIEKAKSLSNHAMKALNKYKRAKVLD---SSRFINE---DOPEKATYQOALNHYS	8382
Qy	524	LTHTNT-----FASTENSKHNLEKTLKALKYMEDYSLRNI-----VV	560
Db	8383	IITHROTENMPDPIVINSITHELETAQNHLHGDKLAHA-----OODAAVINGLITHLNVA	8437
Qy	561	EKEL-----KYKYNL-ISKIEMETIYVE-----NIKQEOGFEEKITKD	600
Db	8438	ONEVINTNTNATTREKVAQKLNDAQALDAMETLLOVVAHKNNIINDOSKYLDNE--SKY	8495
Qy	601	ENKRPDEKILLEYSIDIVKYOVQVLMKNIKIDELKKTOL-----IL-----KNVELK	644
Db	8496	QOOYQDRVADAEQLLNTQNTPTLEPYKVIDIVKONVLANKELILGAEKLSYDKSNADEIK	8555
Qy	645	HNHHPNSYKOBENQ-----EPYLLIVLKKELDK-LKPYMKRVESELNEEKKNIKT	694
Db	8556	HNNTYLNNAQKOSIKIDMTISHAALRTYEQVQLLOQAKILDEAMKSLSEKTOYVITDTLPPYT	8615
Qy	695	EGOSDNSEPTREGETIQOATTKPQOQAGSLEDSDYQAOAOEOKOAPPV--PVEVPEAK	752
Db	8616	EASEDKKRRKQYQYSHAOAI--DKINGSVSLDQYRQALBEOITQASENDODQVBEAK	8673
Qy	753	AOVPPAPVANNKTEVNSKLDYLEKLEYELINTSYICHKTYILVSHSTMEKILKQYKITYKE	812
Db	8674	V-----HANOTIDOTLHNLSTL-----OOOTAKE	8696
Qy	812	ESKSLSCDPLDLEFNIONNIPWYMSFDSLNSLSOLFMLEYEKBYMCNL--YKLNKND	870
Db	8697	---SYKATKLELEIATYSNNA-----QALNKYMKLEQPIINHADSVENSQNTYQADD	8746
Qy	871	KI-----KNLLEAKKVSYV--KTLSSSSMOPL-----SLT	900
Db	8747	KIAVDEALHEGODIOKTNATQNETKOALQOOLIAVETSLNGEERLNHAPRALEYIKSLE	8806
Qy	901	PODKREVSANDTSHSNINLSIKLEPNILSLGKNINYOELIGO-----KSS	948
Db	8807	KINNAQKALSD---KYTQSHDLLELHLYNEBTLN---DIMGELANAIVANNAPTAS	8860
Qy	949	ENFEYEKILKSDTFEYNSFTNFYKSKADDINSIN--DESKRKLEBEDINKLKTKTOLS	1004
Db	8861	INYI-----NADNLRKDNFTQALINNARDALNKYQGNLDNFMAIDPFKDDIFKTDALN-G	8914
Qy	1005	FDLYNKYKLEKLELFDKKKTYVGKYM-----QIKKLT-----LKEODESKL	1046
Db	8915	IERLTAASKREKILDLKFINKAQFTHANDEITMNTNSIAQSRIVYNOAFDNDAMKSLR	8974
Qy	1047	NSLNRPKHVLONFVFNKKKEAEIAETENTLENT-KILKHYKGLVYKNGSESSPLTKL	1105
Db	8975	DELNNQAPPYQASSYNTNSBEDLK-QOFHIALSNMAKYLAKE-----NC-----KNL	9020
Qy	1106	SEESIQO-----TEDNVASLENFVKVSLBEGKLDKNDINLEKKNKLSYLSGLHHLIAELKEV	1160
Db	9021	DEKOLQGLKQVIEPTKALNGIOLRSLAKKKALOYO---OSIASYINAOIRI-----	9069

QY 1161 IKNNKNTGNSPSENNNTDVNNALESYKFLPEGTDAVAVVSESGSDTLEOSQPKPASTHY 1220
DB 9070 -----AENNHNSSDLSLANLSTKASDLDNNAKDL-RDITEEN-----STSV 9111
QY 1221 GAESNTITTSQNVDEVDVY-----IYPIGESEEDYDGLQVYVTEA 1264
DB 9112 PMSVYINADKNLOJEFEDLQASATSSKTSNPATIEIEGLSOAIV-DTKNALNGE- 9169
QY 1265 VMSVYIDNLSTIKENEYEVLYKPLAGVYRSLKOLENNVMPFNVAVKDLSRFRKREN 1324
DB 9170 -----QRLATE-KSKDKLIKGLDKLNAQLE-----DYTN-----KVN 9202
QY 1325 FKNVLESDELIPYKDLTSSNVVYKDPYKFLNKEKDFLSSYNYIKDSI---DTDINFAND 1381
DB 9203 SANLTLE-----LSQLTQSLTELNDKMKLL-----RDKLKTLYNVVKASLVRNADYV---- 9250
QY 1382 VLGYYKILSEKKSDDLSIKKIYINDKQENKEYLPLPNNIEFLYKTVND----- 1430
DB 9251 -----LKRQFNKALKKEAKGVLNKSGTNVN-----INDIQLHLLQIDNAKDQLNGERRL 9299
QY 1431 -----KIDFVHLEAKVLYNTYTE-----KSNVEVKI-----KELN-----YLKTI 1466
DB 9300 KEHQQSEVFILIK-ELDILNNKOKAIIINOIRASKDKIKIINOIVDAIELDNMOGLKEH 9358
QY 1467 QDKLADFKNNNFVGIADLSTDYNH-----NNLTFLSTGYFENLAKTVLSN 1515
DB 9359 VAQLTATTTDN-----TEYLNADBDHKLOYDYAINLANLVLDKENGNTNK-DANIITIGION 9413
QY 1516 LLDGN-LOGMLNISQHCYKQKOPNSGCFRHLDERECCCLLYNKQESDCV-ENPNP 1572
DB 9414 MODARALLNGIERLKDAQ-TKANNDIKDTLRLDEIEHANATSNSKAQAKQVNEEAR 9472
QY 1573 TCNENNGGCDADAKCTEEDSGSN 1595
DB 9473 AUNINDATSNDLVNOAKDEGOS 9495

RESULT 7
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 4.6%; Score 387; DB 4; Length 3696;
Best Local Similarity 20.9%; Pred. No. 3,4e-12;
Matches 314; Conservative 231; Mismatches 592; Indels 368; Gaps 67;

QY 114 SONSSDSADKSYA-----DLKRVNRYLLT-IKELK-----YPOLDTLN 152
DB 2344 ANNKSTNEKSAIDRVNNAKIDAINNTTATTOLVNDAKNSGNTSISQIILPSTAVKN 2403
QY 153 HMLTQDNHGEFYLD-----GYEINELLYKLNFFYD--LLRAKLDVCAANDYCO 202
DB 2404 ALAALASEAKKNAITIDOTPNATAEKEEAPANNKVRLOEADANITIAHTTDEVNIIKNO 2463
QY 203 IPFNLIKIRANELDVILKKLVFGYRKLPLDNIKDNGKMEYIKKNNKTIENINELIEESKKT 262

DB 2464 AVON--INAVQEVIRK-----QNVKN-----OLNOFIDMQKTI 2495
QY 263 IDKNKNTAEKKKKLYQAQYDLSYNNKOLEEHNLSVLEKRIIDTLKKNENKELLD-- 320
DB 2496 IENTPDLTLEKKA-----NRLQO---NVLTSTSELIANVANHNEVDQALDRA 2541
QY 321 --KINEIKNPPANSNGTNTL-----LDKNNKIEEHEK-----EIKEIAITIKF 363
DB 2542 RPKIEEIV-POYSKRDDVNAIQEAFNSQTOEIOENQEAATENKTBALKINOLLNQAKY 2600
QY 364 NIDSLETDLELEYIAREKKNIDISAKYETKESTEPNENYVPLSYNDINNALNEL 423
DB 2601 NIDQ-----AQSNKDVD-SAKTRSJODIDQIO-PHQOTATGRHRLNERANQO 2646
QY 424 NSFGDLINPEDYTKEPSKNITYDNERKKFLINEIKER--KIEKKIESDK----- 472
DB 2647 QS-----TIAHPNSTIEERQEBASAKIQEYLKAIAIKIDGQINDVEKTVVNGIAE 2698
QY 473 -----SEYDRSKSLNDITKEYEKRLNEIYDSKFNNNIDLTNPEKMGKRSYKVERL 524
DB 2699 IENILPATYVKKAKADVNAKEQKNL-----QINSDEATEEKKLVASDLNLHVENF 2751
QY 525 THHNTASTENSKHNLEKTLKALKYMEDYSLRNIVYEKELKYKNLSKIEEIE----- 579
DB 2752 TNOAIEDAPDTNOVNER-KNGIGITRD--IOPLYVK-----PTAKSKIESAVEKKTE 2803
QY 580 -TIVENIKKDE-----EOL--FEKRTIKDEN-----KPD--EK 607
DB 2804 INOTQATTHDEYREGINOLNQHEKAKNDVNSQTOQVBNMAQNSLDOINFRPFSK 2863
QY 608 ILEVSIDIVVQOVKVLNMLNKIDELKKTOLILKNVELKNI-----HVPNSYKQENK 658
DB 2864 RANVAEIVKAQ-----QKKIDIE--OESATQEEKDNALQHLDBQVKEIINSQANT 2915
QY 659 QEPYLYIYLKKEIDKLK------FMRVESLINEEKN--IKTEGOSDNSEPTGE 708
DB 2916 D-----NEVDNAKTSGLNNTITEYREY-----NKRKNALIKLYDVSQTEALING- 2960
QY 709 ITGQATTPGQOQSGALEGSGVOAQOQ-----KQAPPVVPVPEAKAQPVPAPVNNK 765
DB 2961 -YPDATEDEQLQEANSKL--NKILDAKKQIGLAHTNNEVDYDINEVSQKMKITLPRVDIK 3017
QY 766 TENVSKLDYER--IYEFINTSYICHK-----YLVSHSTNEKILKQYITKEESK 816
DB 3018 AVARSVLNALAKOLIKTFEFTADVTHEERNDAINHKEOLSLVFANIEKDRKIOVAODE 3077
QY 817 LSCDPLDLIF-NIQNN-----IPVYMPDSINNSLSQLEMEIYERKMCNLYKLC----- 867
DB 3078 LGLNELNLSIFINITQKPTARKAISGMASQOLNNSINNTPYATEERQOLA-LNKVKAIVDD 3136
QY 868 DNDKIKNLEAKKSVTSYKTLSSSMOPLSLPQOKPEVANSADDPSSHNTLNSKLFE 927
DB 3137 ANEKIREANTDSEVLGT--KTNAITLLQASADVQKPPA-----FE 3176
QY 928 NISLGRKNRYIOELIGQKSSSENFYER--ILKSDTFYNESTFNKYKSKADINSINDS 985
DB 3177 EI---NAQAEIQERIRINGNSDATREKEEKALQVDVLVHNSFTT-----INNVKNQ 3225
QY 986 K-----KKLEEDINKLKTLOLSFDLYNKYKLELRFD-----KKTYGKYMQIKKYL 1036
DB 3226 EYVDTDKTIEALHKIKPSTIKPQALNETTQLOLQORDIKNNKSTVEEKASAIKDL 3285
QY 1037 LKEQESKLSLNNKRVYONFSVFNNKKEA--EIAETENTLE-----NFKIILKH-- 1087
DB 3286 KTAARIAELDKQOTNEEYK-----NIKQSIDETSKLPIETELKSARNIEHKAERY 3338
QY 1088 YKGLVYVYNGESSPLKLTSEESIQEDBNYASLENFVL-SKLEGLKLDMLNEKKKLSYL 1146
DB 3339 IRLINDNEATKEEKDIALINQDITLTQANVSIDQALNEAVANRAKELANSINKISYI 3398
QY 1147 SSGLHLIAELKQVINKKNTGNSPSENNNTDVNNALESYKFLPEGTDAVAVVSESGSDT 1206

Db 3399 AIRKEFAIAEIOELADK-----LNKFKO-----SQEAT-IEBKQAI 3435
QY 1207 LEQOPPKPASTHVGASNTITTSQNDVDEVYIIVIFEESE-----EDVDDLGCV 1259
Db 3436 NELEOALKSATNHHSOSNNESVSAALKESTLSIDSEIOAHKKLEAKAYIDGSD---- 3491
QY 1260 VTGEAVPTSVINDITLSKIENEYEVLYLPLAGVYRSKLKOLEN-----NVMTFVNVK- 1312
Db 3492 -----DKINDISSRATNEKOIFVSKLALINRHKQIDEAETVSVETIVRNEKVE 3543
QY 1313 -DILNSPKNKRNKNVLESOLIPYKDLTSSNYV-----VKOPYKFLNKKRDKFLSSYNY 1367
Db 3544 ADKNLSYVRKAKASKETELEDVHKOMINANLASTFVKONARTLLINEIVSNALSQLNK 3603
QY 1368 IKOSIDIDINAFANDVLYGYKILSEKYSDDPSIKKIYINDKQGENEKYLPFLNIETLYKT 1427
Db 3604 VTTKKEVD-ELVNETIEKLSIOIREDKILSSOSSTSMTEKSNOCYSSENNITKSLPEA 3662
QY 1428 VN-DK 1431
Db 3663 GNADK 3667

RESULT 8
US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
APPLICANT: RATTNER, JEROME B.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY EXPRESSED KINETOCHORE PROTEIN,
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DOFFMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-1

Query Match 4.2%; Score 356.5; DB 1; Length 3248;
Best Local Similarity 18.9%; Pred. No. 1.2e-10;
Matches 350; Conservative 304; Mismatches 631; Indels 569; Gaps 79;

QY 125 YADLKHR--VANNYLLTTELKYLPOFLDTNMLTCDNIHGFYKLIOS---YEETINELLY 179
Db 709 FSDQKHQKEIENMCLKTSQLT-GQVEDL-EHKLOLLSN-----ELMDKDRYQDLHAHYE 761
QY 180 KLNFFDLLRAKLANDVCAND-----YCOIPEFLKIRANELDYLLKLVFGYKRPIDNI 231
Db 762 SLR---DLKSKDASIVYNEDHQSILAFDOQPAHHHSFAN-----IIGEGSMPSE 810
QY 232 KDNYGKMDYIKKKKTTEINENLIEBSKRTIDKNKNAKTREEKKKLYQAOYDLSIYKQ 291
Db 811 RSECRLEADOSPKNKSAIIONNVDSLEFLES-QKOMNSDLOKQCELVQINGEILENMK 869
QY 292 LEEAH-NLISVLEKRIDTLKKN-----ENIKELDKINEIKNPAPSGMTPTLL 341
Db 870 AEOHQSVAETSGRISLQEDTSAHQVVAETISALENKKKELO-----LL 916
QY 342 DKNKIEHEKEIKEIAKTIFNIDS-----LTPDPLEYLYLRKKNKIDISAKVETKE 396
Db 917 --NDKVEEQAEIOELKKNHLEDSLKEIOLLESTLSL--KKEMSIIISLN-KREIEE 971
QY 397 STEPEYNGVYTPYSTYNDINNALNELNSFGDLLNPFYTKEPKSNITDNERKKFIMEI 456
Db 972 LTOENG-----TLKEINASLNOEKM-NLIO-----KSEFANTID-EREKSISEL 1014
QY 457 KEKIKIEK---KRIESDKSYEDRS--KSLNDITKEKLEINEIYDSKFNNDITLTF 509
Db 1015 SDQYKQEKLIILQCEIEGNAVIEDLSQYKKAQEKNSKLEBLLDCTSLCENRKNLEQL 1074
QY 510 EKMKGKRYSYVEKLTHTNFTASTYNSK--HNLEKTLAKL-YMBDSLRNIVVEKELKY 566
Db 1075 KEAPAKEHQEFLTKL---AAEERONMLLELEVQALREMTDNONNKSSEAGLQK 1130
QY 567 YKNLSKIENIEFTLYENIKDEQOLPEKRTKQE-----NKRDEKILEVSDI- 614
Db 1131 EIMTLKEBONKQKEVANDLOENEDOLMKVMTKHECONLESEPIRNSYKERESEKNOCNF 1190
QY 615 --KVQYQAVLNMKIDELKKTOLILKNEVK-----HNHV----- 649
Db 1191 KPQMDLEVKEISLDSYNQVLQVLEAMLNKELKIQESEKEKQLOHQLTIRGDLSTNL 1250
QY 650 -----PNKYQKQENOEPIYLLVKKKEIDKLVPMPKVS 684
Db 1251 QDMOSQEIISGLKDEIDAEEYISGPHELSQDNDAHLOCSLOTYNNKLN-LEKICEI 1309
QY 685 INEKNKITEGQSDNSPSTEGEITGOATKPPGOAG-----SALEGDSQV 731
Db 1310 LOAEKVELYTELINDSRSCIT-----ATKMAEYVKLNEKYLINDSDGLHGLHYE 1362
QY 732 ---AQAOEQKQAPVPVPPVPEAKAQVPTPAPVNNKTENVSKLDYLEKLYEFLNTSYI 787
Db 1363 DIPGEFGEQENQHPVSL-----APLDEBSNS-----YEHLTLS- 1396
QY 788 CHKYLIVSHSTNNEKIL---KQYKTKREEBSKSCDPLDLFNINQNNIPMYSMFDSL- 843
Db 1397 -DKEVOMHFAELOKRFSLIOSEHKILHDQHCOMS-----KMSLQTYVYDSLK 1443
QY 844 -NNSLSQLFMEIYKEKEMVCNLYKLDKNDKIKNLEAKVSTSVKTLSSSMQ----- 895
Db 1444 AENLVSTNLNLFQGDLYKEQOLGEBGLVPSLSSCCYVDSLSLSDSSFYRLLEQT 1503
QY 896 -PLSLTPQDKPEVSAN-----DDTSHSTNLNLSKLFENILSLG 933
Db 1504 GDMSLSLNLEGAVSANQSVDEYFCSSLOTYVDSLAKENLVSTNLNLFQGDLYKEQOLG 1563
QY 934 KKNKIYQEL-----IQKSSSENFYEKIL---DSGTFYV-ESTTNVYKSKADI- 978
Db 1564 LEEGLVPSLSSCYVDSLSLSDSPFYRLLEQTDGMSLNLNLEGVASANQSVDEYF 1623
QY 979 -NSLNDSESKRKK-----LEEDINKLTKTLOLSPDYLYKKYKIKLERL 1018
Db 1624 CSSLOEENLTKETPPSAPAKAVEELESCEVYRQSLERLEKEMBSQGITMKKKEIOELQL 1683
QY 1019 FD-----KKTYVGKYYKKQIKKTLFLKLEOLSKLNS-----LNNPKHYLON 1058

```

Db 1684 LSSEROLDCLRRQYLSNEMOQKLTSTVLEMSKLAERKQOTOLESELEVARLOLG 1743
QY 1059 FSIV-----FFKKKEAELEENTLENTKILKKH----- 1087
Db 1744 LDLSRSLGIDTEDAIOGRNESCDSK-EHTSETTERPKHVOICDKDAQOQDLNDI 1802
QY 1088 ----XKGLVK-----YNGESSP-----LKTLS----- 1107
Db 1803 EKITEGAKPKPGEGSGEDSPDINTEPPEDKQSGSSECSISLSTFGNALVPMDFLGN 1862
QY 1108 ---ESTQEDNYSLENFVLSKLEG---KLDNLLEKKKLSYSSGLH---HIAEL 1157
Db 1863 EDIHNLQLVKETSNEMLRLHLHYIEDRDKRVESILN---EMKELDSKLHLEVOQMTKI 1918
QY 1158 KEVKKKNYTGSPSENNTDVNNALLESY----- 1185
Db 1919 EACIELEKTVGLKKE-NSDLSKLEYFSCDHOELLQRYETSEGLNDELMHADKSSRD 1977
QY 1186 -----KKPLPGTDVATVSVSGSDTLLEQSPKKRPASTHGAESNTITTSO- 1231
Db 1978 IGDNNAKVDSMKERFLDVENELSRISSEKASIEHE-----ALYLEADLEVOQTEKL 2029
QY 1232 --NWDEVDVITVPIFGESSEDDYDLGOVVT-----GEAVTPS-----VIDNLSKIE 1278
Db 2030 CLEKDNENKQKVIYCL-----EDEL-----SVATSERNQLRGELDTMSKTTALDQLEKKK 2081
QY 1279 -----NEYELVLYKPLAGVYRSLLKQLENNVMTFNVNDKILNSFRNKRENFNVLE 1330
Db 2082 EKTQELESHQSECHICIQVAEAEVKEKTEL---LQTLSDVSELKDKTHLOEKLOS-LE 2137
QY 1331 SLLIPYKDLTSSNYVVKDPYKFLNKKR-----DKFLSSYNYIKDSIDTDINFAND 1381
Db 2138 KD---SQALSTLKCELENOIAOLNKEKELVKESESLQARLESSEYDEKILNVSALBAALY 2194
QY 1382 VLGYYKI-LS-----EKXSDLS-----IKKYNKQGENEYKLEPFLNN 1420
Db 2195 EKGEFRLRLSTQOEYHQLRGIEKLVRIEADKKOLHIAELKEREENDSLDKVEN 2254
QY 1421 IETLKYTVNDKIDLEFVILHEAKVLTNTYKSNVE-----VKIKEL-----NYLKT 1465
Db 2255 LRRELQMSSENGEVLILDAENSAEVEITLKTQIEEMARSLIKIFEDLVTLRSEKEMULTQ 2314
QY 1466 IODKLADFKNNNFVGIADLSTDYNNHNLITKFLSTGMVFNELAKTVLSNLDGNLQGM 1525
Db 2315 IOEKQQLSE-----LDKLLS-----FKSLLEEKQOAEI 2344
QY 1526 NTSOHQCVAKKQCPNCGFRHLDERECKCLNTYKQSGKVENP-NPTCNENN 1578
Db 2345 QIKESKTAIVEMLQN-----QLKELNEAVALCGDOEIMKATEQSDPPIEBEH 2393

```

```

RESULT 9
PCT-US95-16216-1
Sequence 1, Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
APPLICANT: Ratner, Jerome B.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 4.2%; Score 356.5; DB 5; Length 3248;
Best Local Similarity 18.9%; Pred. No. 1.2e-10;
Matches 350; Conservative 304; Mismatches 631; Indels 569; Gaps 79;

QY 125 YADLKRR--VRNLLTIKELKYPOLFQDLNHNMLTCLDNIHGFYLLDG---YEEINELLY 179
Db 709 FSDQKHQKEIEMNCLTSOLF-GQVEDL-EHKLOLSN-----EIMDKRCIYDLHAEBE 761
QY 180 KLNFEYDLRAKLNDVCAKD-----YQAIPENLKIRANELDVLKLVFGYKPLDNI 231
Db 762 SLR---DLLKSDASLVNEDHQRSLAFDQOPAMHHSFAN-----IIGQSGMPSE 810
QY 232 KDNVGMEDYIKKKNKTININELIESKKTIDKNNNAKREEKKKLYQAQDLSYNNQ 291
Db 811 RESCRLEADQSPKNSAIIQNRVDSLEFLES-OKQNSDQKQCEBLVQIKGEIEENLMK 869
QY 292 LEEAH-NLISVLEKRIQDTLTKN-----ENIKELDKINIKKPNPANGTNPFTLL 341
Db 870 AEGMHQFVAEISQKSLQEDTSAHQNVVAETLSALENKEKELO-----LL 916
QY 342 DKKKILIEHEKEIKELAKTIKFNIDS-----LETDPLEEYVLRKNNKIDISAKVETE 396
Db 917 --NDKVETQAEIQELKKSNNHLLDSIKELQLSFTLSLE--KKEWSIISLN-KREIEB 971
QY 397 STEPNFEPNGVYPLSYNDINNALNELNSFGDLINFPDYTEKSKNIYTDNEKKRIINEI 456
Db 972 LQENG-----TLKEINSLNDEKM-NLIQ-----KSESFANYID-EREKSISEL 1014
QY 457 KERIKTEK---KRIEDSKSYEDRS---KSLNDITREYKELLENLEYDSKFNNDIDLPNF 509
Db 1015 SDQYKQEKLLILORCETGNAVEDLSQKKAQEKSKLECLINETSILCENKKNLEQD 1074
QY 510 EKMGRKYSYKVEKLTHTNTPASYENSK--HNEKLTALK-YMEDYSLRNIVYKELKY 566
Db 1075 KEAFKEHOEFITKL--AFAEERNQNLMELETFVQALRSEMTDNQNNKSEAGLQK 1130
QY 567 YKNLISKIENETIETVENIKKDEQOLFEEKITPDE-----KKPDEKILEVSDI-- 614
Db 1131 EIMTLKEEONKQKQEVNDLQENQIMKYKTHQCONLESEPIRNSVRESERNQCNF 1190
QY 615 ---VKVOQVLLMKNKIDELKKTOLILKVNELK-----NHNIV----- 649
Db 1191 KQPMDELVKEIISLDSYNAQVLEAMLRNKKELKLOESEKEKECTQHELOTINGDLETSM 1250
QY 650 -----PNSYKQENKQEPYVLLVKKELDKLVKFMFVRESL 684
Db 1251 QMWQSOEISGLKQCEIDAEKYSIGPHELSTSONDNAHLQCSIQTMKNKLN-LEKICEI 1309
QY 685 INEKKNITEGSDNSPSTEBEITGOATTKPGQAG-----SALBDSQVQ 731

```

Db 1310 LOAEKVELTTELNDSEKICIT-----ATRKMAEEVGLNEVKILNDSEGLHCEIVE 1362
Qy 732 ----AQAOBQOQAPVPPVPEAKAQAQVTPAPVNNKNTENNSKIDYLEKLEFPLNTSTI 787
Db 1363 DIPGEFGEQPEQHPHVSIL-----APLDESNS-----YEHITLS-- 1396
Qy 788 CHKYLIVSHSTNNEKIL--KQYKITKEEESKSSCDPLDLFNIONNIPVYMSFDSL- 843
Db 1397 -DKEVOMHFAELOEKFLSLOSEHKILHDOHCOWS-----KMSLOQTYVDSLK 1443
Qy 844 -NNSLSQLEMEIYERKEMVCNLYKLDNDKIKNLEAKKVSTSVTLSSSMQ----- 895
Db 1444 AENVLSTNLNRFQGDLYKEMQLGLEGLVPSLSSSCVPSDSSLSLGDSPFRALLEQT 1503
Qy 896 -PLSLTPQDKPEVSN-----DDTSHSTNLNLSKLFPENLISLG 933
Db 1504 GDMSTLSLNEGAVSANQCSVDEVCSSLOQTYVDSLKAENVLSTNLNRFQGDLYKEMQLG 1563
Qy 934 KKKNIYOEL-----IGQSSSENPYEKILK--DSDTFYN--ESFTNPFKSKADIT- 978
Db 1564 LEEGLVPSLSSSCVPSDSSLSLGDSPFRALLEQTGDMSTLSLNEGAVSANQCSVDEVF 1623
Qy 979 -NSLDESKRKK-----LEEDINKIKTLQSLFIDLNNKYKTLKEBL 1018
Db 1624 CSSLOEENLTKRETPSAPAKGVEELESICEVYRQSLKEEKMESOGIKMKKEIOLEBOL 1683
Qy 1019 FD-----KKTVGKYKMKIKRUTLLKEOLESKLNS-----LNNPKHYLON 1058
Db 1684 LSSEROELDLCKOYLSEBOMQOKLTSTVLEMEKSLAEKKOTBOLSLLEVARLOLOG 1743
Qy 1059 FSV-----FPNKKKEAIEAETENTLENTKILKKH----- 1087
Db 1744 LDLSRSLIGIDTEBAIGCRNESCISK- EHTSETTERPRKIDVIOICDKDAQOOLNDI 1802
Qy 1088 ---YKGLVK---YNGESSP-----LKTLS- 1107
Db 1803 EKITETGAVKPTGECSEQSPDTNYPEPGEKTOGSSSECISELSPFNALVPMDFLGNQ 1862
Qy 1108 ---ESIQEDNANSLJENKVLKSLG---KLDNINLEKKKLSYSSGHL---HLIAEL 1157
Db 1863 EDIHLQOLRWKETSSENRRLHVIDRDRKVESLILN---EMKELDSKLHLOEVOQLMTKI 1918
Qy 1158 KEVINKNVYTGSPENNTDVNNALESY----- 1185
Db 1919 EACIELEKIVGELKE-NSDLSERLEFSCHOELLQVETSEGLNSDLEMAHAKSSRED 1977
Qy 1186 -----KKLEPGTDVATVSESGSDLEQSOQPKKPASTHVGAESNTITTSO- 1231
Db 1978 IGDNAVAKVNDSWKERFLDVENELSRIRSEKASIEHE-----ALYLEADLEVYQTEKL 2029
Qy 1232 --NUDEVDVIVIPFEESEEDYDLOQVY-----GEAVTFS---VIDNLSKIE 1278
Db 2030 CLEKONENKOKATVCL---EBEL---SVYTSERNOLRGELDTJMSKTTALDOLSEKKK 2081
Qy 1279 -----NEVEVLYLKPLAGYRSLKQLENNVMTFNNVNDIINSRKNREKNKLYE 1330
Db 2082 EKTQLESHQSEBCLHCIOVAEAEVKEKTEL---LOTLSSDVSELLKDKTHLOEKLOS-LE 2137
Qy 1331 SDLIPYKDLTSSNVVKKPYKFLNKKR-----DKFLSSVYVYIKDSIDTIDINPAND 1381
Db 2138 KD---SQALSTLTKCELENOIQAOLNKEKELLYESESLOARLESQSEYKLVNSKALEALV 2194
Qy 1382 VLGYYKITLS-----EKYKSDLS-----IKYINDKOGGENKTYPLFLNN 1420
Db 2195 EKGEFALSLSTQEEVHOLRGIEKLRVRIEADKKOLHIEKLEKREERENDSLDKVYEN 2254
Qy 1421 IETLYKTYNDKIDLFVILHEAKVLNITYEKSNE-----VKTKEL-----NYLKT 1465
Db 2255 LERELQMBEENQELVILAEKSKAEVETLKTOIEBMAKSLIFEIDLTYLSEKENLTKO 2314
Qy 1466 IODKLADFEKKNNNFVGIADLSTDVYNNHLTKFLSTGVFENLAKTVLNLDLGNLQML 1525

Db 2315 IOEKGOLSE-----LDKLLSS-----FKSLLEKEQOAEI 2344
Qy 1526 NISQHCQVAKQCPONSQGRHLDERECECKCLINTQESGKYENP-NPNCENN 1578
Db 2345 QIKESKTAIVEMLOM-----OLKEINAEVAALCGOELMKATQESLDLPIEEBH 2393
RESULT 10
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCF
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2
Query Match 4.2%; Score 353.5; DB 4; Length 2285;
Best Local Similarity 20.1%; Pred. No. 1.2e-10;
Matches 311; Conservative 244; Mismatches 546; Indels 443; Gaps 73;
Qy 119 DSDAKSYADLKHVRNRYLLTTEIKEL-YPQLPFLTNHMLTLCNI--HGFXYLIGYEI 174
Db 874 DSAKTFEDASKETIK---SKDELKQYQIADYNDKGRKMDLIADDDYKYAAD---KA 927
Qy 175 NELLVKNFDFDLRAKILND---VCANDYCOIPFNKIRANELDYKLKLVGYSRRPLDNI 231
Db 928 KQSMKKAQSDIESGNAKVDSVLSIANVYSIDISNTLKTISDVVYNNK-----NL 978
Qy 232 KQNVGKMEDYIKKNNKTJENINELIEBSKTTIDKNNKNTKEEKKKLYOAYQDLSYKQ 291
Db 979 KQDLDPPE-----LEKFSSLSGKLOEKMOKALDSDEKAFDAKDL--QSLLETYSK- 1029
Qy 292 LBEAHLISVLEKRIIDTLKKENIKELDKINEIKNPSPANGNTPNPLDNNKIEIEE 351
Db 1030 ---SSSIDVFKASPD--KAQKNIKDGKSLSSVK-----SEVGDIGETIAGNGNAEDG 1080
Qy 352 KEIKEIAKTIKENIDSLFTDPLELEYLREKKNKIDISAKVETKESTEPENYNGVTPPL 411
Db 1081 KKLKEAL-----DANSVDIKAAI--KMSDAMQF-DSVQDVL 1115
Qy 412 SYNDINNALNEINSFGDLINPDYTKEPSKN-----ITYDNERKKFINEIKETIEKK 466
Db 1116 NQDIFNNTKQVAPLNDLEKMAEKGKSIISANANTLQIKDKELAQAI SIENGCVINRDE 1175
Qy 467 IESDKSYEDRSKINDITKEVEKLLNELYDSKFPNNNIDLTPEFKMMGGRYSYKELTH 526
Db 1176 VTKQKRVKID--AYNDWYTSNKLAK--TEVNNAKITTLNADTL-----RIDSL- 1219
Qy 527 HNTFASYENSKHLEKTLKALYMEDYSLRNIV-VEKELKY-----YKNLSIKIE 575
Db 1220 -----KILRKEKKLMSSEKELSDLEVKSIINNVADAKKELKREKMLDQGVGSN--SQIE 1272
Qy 576 --NEIETLVENIKKDEBOLFEKKTITKDEKDPDEKTLLEVSDIYKVOYOVAYLLMKNIDELK 633
Db 1273 AMQSVKSALESYISASE--EATSTQEMNK--QALVLEAG-----TSLENNMTDOOEK 1318
Qy 634 TQULIKNVELKNIHVPVSYKOE-----NKQ-----EPYLLIVLKKKEIDLK 675
Db 1319 A-----NEETKTSMTVVDKYEALAEKVNAEIDKYKQVNDPKYSOKRDAIKKIKALQ 1373
Qy 676 VFMPVRESLINEKKNIKITEGOSDN-----SEPSTEGETTGOATTGPOOQ 720

```

Db 1374 ----OKKKLMQDAKLLDKQIKSGNITQYGVTTSSGGTSPSTGSGYSKGYSYINSA 1429
Qy 721 ACS-----ALEGDSVOAOEOEKOAPVPVPEAKAOPVPPAP-----VNNKTEYNS 770
Db 1430 ASKYNVDPPLIAVIOEGSFGNAKASGV-----GAMGLQMLPALAKSLGVNNAIDPIQ 1484
Qy 771 KLDYLEKLEYPLNTSYICRKYILVSHSTNEKILKOYKITTKEESKLSGCDPLDLLENIO 830
Db 1485 NV-----MGCTKYLAQO--LEKFGNVEKLAALVYNAAGPQVNIKYGGIIPPEKTONYV 1534
Qy 831 NNIPVWYMSFDS-----LNNLSQLFMEIYEKENWCNLYKIKLNDKXINLLEAK 880
Db 1535 KRTIMANYSLSLSSANTSSIASYTTNNSAFRYSKYGOEGSLRSSPKHTD-----FAA 1587
Qy 881 KYSTSVKLTSSSSMOPLS-----LTPQD-----KREVSAND-----911
Db 1588 KAGTAKLSQSGKVOIAGYSKTAGNWNVIKODGTAVAKYMHMLNTPSVAKAGSVKAGOTI 1647
Qy 912 ----DTSHTNLLNSLKFENILSLGKNKNTIOELIGOKSSENFYEKILKSDPTFYNEF 967
Db 1648 GRVSTGSGNTHLQIDQNGKTIIDPEK--YMOGIGTS-----ISDAQAEAEBO 1696
Qy 968 TNPVSKAD-----DINSUND-----ESKRRKLEEDINKLKTQLSFDLYNKY 1011
Db 1697 QGTAQAKSLLSGDISSYNDQIOELQYELVOSKLDPEDKRIGDDVRIAKDESMAKRY 1756
Qy 1012 ----KLLERLPDKKKTGVRY-KMOIKLTLLEKQESKLSLNPNKHLNPNFVFNKK 1066
Db 1757 TSDSKFRYTSQKKAVALBOAKIOQOKVMIQKEIKTN-KALNS-----1800
Qy 1067 KKEALTEENTENTLKIILKHKGVLKYNGESSPLKTISEESIQF--EDNVASLENFY 1124
Db 1801 --AQRQLOEELKQAKLIDISYQDQY-----RELQKOLVOSKVBETLKSIE--KS 1846
Qy 1125 LSKLEKLDKNDLNL-----EKKKLSYSSGLHLIAELKEVYKNNKYTNSPSENNTDV 1178
Db 1847 SSKTQOKIKDVNKKIMTEDEDEKVKYIKOILIOQOQKEA---KKYI-----1892
Qy 1179 NNALESYKFLPEGTDAVATVSESGDTELOSQPKRPASTHVGAE-SNTITTSQNVDEY 1237
Db 1893 -KQLEBQKKAQGFPIQIOQIT-----EEMQNKQKQKDFMLEYNTKSKIKIDIKYSL 1944
Qy 1238 DQVITYPIGESEEDYDQGVVYTGAVPVSVIDNLISKIENYEVLKPLAGVRSUK 1297
Db 1945 ADE-VVSITKEMEKMRDI-ELEAHQKATQDLDEI-DKTDE-----AKFOKELK 1992
Qy 1298 KOLEN-NVMTFNVNVKDIINSFKNREKNVLESLLIPYKDLTSSNYYKDKYKFLNKE 1356
Db 1993 ERQDSIQKLTDOINOYSLDSEBGRSK-----VKELTEQ-----LOKE 2030
Qy 1357 KRDKFLSSYNYIKDSIDTDINFANDVLYGYYKILSEKYSDL-DSIKKYINDKQGENEKL 1415
Db 2031 Q-----LDLDDFLDKR-----ESNKRKEALQDLEK--DEESINNKYD 2066
Qy 1416 PLPLNIEITLKYVNDKIDLFVHLEAKVLNITYEKSNEVEKIKELNYLKTIOKLADEFK 1475
Db 2067 NLVND-ERAFKLEKDI-----MNGKITDIADK 2092
Qy 1476 N-NNEVGIAIDLSTYNNHNLTKFELSGWFEENLAKTVLSNLD 1518
Db 2093 QLNELF-----SKFINTNM--ESIGKSISNNLID 2118

```

RESULT 11

```

; US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRILLHE, PIERRE
; APPLICANT: DUBBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B

```

```

; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

```

Query Match 4.2% Score 351; DB 4; Length 1786;
Best Local Similarity 20.0% Pred. No. 1.2e-10;
Matches 269; Conservative 243; Mismatches 437; Indels 398; Gaps 68;

```

```

Qy 229 DNIRKDVGMKEDYIKRKNKTIENINELIEES-KTIDKNKNAKREKKKLYQAQYDLSI 287
Db 757 ESVEENY--EESVAENVE--ESVAENVEESVAENVEESVAPEVEIAPSVESVAPSV 811
Qy 288 YNKQLEE-----AHNLISYLERKIDPLKKNENIKE-LLOKINKKPNPANSNGTFTL 341
Db 812 EESVAENVATNLSDNLNLGIGIET---ETKDSILNIEEVEK-----NVVTTIL 860
Qy 342 DKNKKIKIEHEKE-----IKELAKTIKFNIDSLFTPDLELYLREKKNKIDISAKVETRES 397
Db 861 E--NVEETLAEVSFTFSNLEIEIQEN--TITNDTIEEK--LELHEHNV-TSAALENQTS 912
Qy 398 TEPNEYPNGVITYPLSTINDINNALNELNS--FGLLNPFDTKPSKNITYIDNEKKAFFINE 455
Db 913 EEEKK-----EVDVIEEVEKEEVAATTLIETVEQAEKKSANT-----ITE 951
Qy 456 IKKIKIEKKIKESDKSYSDRSK-----SLNDITKEEKLNLNIEYSKFNNDITLTFEK 511
Db 952 IFE--NLEENAVESNEVAENLEKLNETHVNTYVDKVEEVEIYSGESLENEMDKAFSE 1009
Qy 512 MNGKRYSYKVELJHNTFASSENSKHNLEKLYKALKYMEDYSLRNIVKELKYYKNLI 571
Db 1010 I-----FNVKGIQENL--LGMFRSIE--TSIVIGSEKVDLN--1044
Qy 572 SKIENEITLVENIKKDEQOLFEE--KITKDEKKPDEKILEVDIYKVQY-----QKV 622
Db 1045 ---ENVVSSILDNIEKKKLELNLKLENISSTEGVETHEVBONYVVDVADPAMQDQL 1101
Qy 623 LKNNKIDELKKQOLIKNVGLKHNHVPNSYKQEN-----KQEPYULIVLKEIDKL 674
Db 1102 GILNEAGLKEMFNLEDV-----FKSESQVITYEETIKDEPVQKEVEKEVYSII 1150
Qy 675 KVPMPRYVESLINEKKNIKIKEGSDNSEPSTEGEITGQATTKPGQAGSALGEDSVQAQA 734
Db 1151 EEMEEENIVDLLEKEKDL--TKDMIDAVEESIE-----ISSDS--KEET 1190
Qy 735 QEOQKQAPVPVPPVPEAKAOPVPPAPVNNKNTENSVKLDYLEKLEYPLNTSYICHTIYV 794
Db 1191 ESIRKDEKDVSLVVEEVOQ-----NDMDESV-----EKVLELKN-----1224
Qy 795 SHSTNMEKILKOYKITTKEESKLSGCDPLDLNLNIGNNIPVMSMPSLSLNSLSOLFMEI 854
Db 1225 ---MEELMK-----DAVE-----INDITSKILEET 1247
Qy 855 YE-KEMVCNLYKLDNDKIKNLLEAKKAVTSYKLTSSSMQPLSLTPQDKPEVSANDPT 913
Db 1248 QELINEVADL--TKDMKELKEL-----EKALSSED-----KEIIDAADPT 1285
Qy 914 -----SH--STNLSNLSLKFENILSLGKNKNIYQELIGOKSS-ENFYEKILKSDPTFY 963
Db 1286 LEKVIEEHHITTTLDEYVEL-----KQVEEDKIEKVSDDLMLDEDLIKE-----1330
Qy 964 NESFTNVSXKADINSLNDESKRKKLEEDINKLKTQLSFDLYNKYKIKLERLDPDKK 1023

```

```

Db 1391 -----VKEIKELSEI--LEDYKELTIEIDLIEKK--EIEKHDEKFEFEAEIKDLEA 1382
QY 1024 TWGKKMKQIKKILTKLEDESILNPNKHYLFQSVFNNKKRAELAEFNTLENKI 1083
Db 1383 DI-----LKEVSSLEVEEKLEVEHKE-----EVEHITSGD-- 1416
QY 1084 LKHVKGVLKYYNGESSPLTKLSESIOTEDNVALENFK--VLSKLEGLKDONILEKK 1141
Db 1417 --AHIKGL-----EEDDLLEVDLKGSLDMLKQDMLG--DMOKE 1453
QY 1142 KLVYSSGLHILIAELKVIKNNYNTGSPSENNNDVNNALSEYKFLPEGTDAVAVSE 1201
Db 1454 SLEDVTKLGERVESLJKDLVSA--LGMDEQMKTRKKAQRKLEVL-----LKEEYVE 1506
QY 1202 SGSDTLEBS-----QPKKPASTHGAESNTTTTQNDVDEVDVYIPIGESEEDD- 1254
Db 1507 EPKKTKTKKVPKIDKEPEDEIVEEMKODEIEDEVEDIEEDKVEDIDEDIDE 1566
QY 1255 DLGQVVTGEAVTPSYIDNLSKIENEYEVLYLPLAGYRSLSKOLENNVTFFNVKDI 1314
Db 1567 DIGE-----DKDEVIDLIVOK--EKRIEKVAKK-----KLEKVEBESVGLKHHDEV 1614
QY 1315 LN--SREKNR--ENKKNVLESDLIPYKDLTSSNVYKDPYKFLNKKRDKFLSYN-YIK 1369
Db 1615 MKYQKIDKEVDKESVKALESK-----NDVTN---VLKQNDPFSKVK--NFKYKXVFPA 1665
QY 1370 DSIDTDFADVIGY-----KILSEKYSKSDLSIKKYINDKOGENEKTYLPFLNN 1420
Db 1666 PFISAVAFASVYGFESFSSCVTIASSYTL--LSKVKTIN----KKKEPPFSYF 1718
QY 1421 IETIKTYNDKIDLEVIHLKAVLTYEKSNEVEYKIKELVYKTIODKADFKNNNV 1480
Db 1719 VFDIFKNL-----HLOQMKERFSK--EKNNNYI 1746
QY 1481 GIADLSTDYNNHNLTKELSTGMVEEN 1507
Db 1747 EVTNKAEKKGNVQVNTKTEKTKVDKN 1773

RESULT 12
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001

```

```

; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 4.2%; Score 350.5; DB 1; Length 2482;
Best Local Similarity 19.8%; Pred. No. 1.8e-10;
Matches 355; Conservative 281; Mismatches 617; Indels 537; Gaps 79;

QY 125 VADLKHR--VRNYLTITLIELKYPOLFNLNHLITLQDNIHGKRYLLDG---YEELINELLY 179
Db 77 FSDQKHQKEIEMWCKLTSLT--GQVEDL--EHLQLLSN-----ETMDQRCYQDLHAEYE 129
QY 180 KLNFFYFDLIRAKLINDVCAND-----YCOITPFIKIRANELDVLKLVFGYRKPPLDNI 231
Db 130 SLR---DLKSDASLVYNNEDHQBSLAFDQOPAHNHSPAN-----IIGQGSMBSE 178
QY 232 KDNVGMEDYIKKNNKKTIEINELIEESKKTIDKKNATKEEKKLYQAOYDLSIYNQ 291
Db 179 RSECRLEADQSPKNSAILQNRVDSLEFLES--OKOMNSDLOKCEBELVQIKGEIEENLMK 237
QY 292 LEEAH-NLISVLEKRIIDLKNN-----ENIKELDKINEIKNPPRANGTPTNLL 341
Db 238 AEOMHQSFAVETSQRIKQIEDTSAHONVVAETLSALENKEKELO-----LL 284
QY 342 DNKKIEEHEKEIKELIATKIFNIDS-----LFTDPLELEYLRKNNKIDISAKETKE 396
Db 285 --NDKVEYEOAEIOLKSNHLLIEDSLKELQILSTLSIE--KKMSSTISLN--KREIFE 339
QY 397 STEPNEYPNGVTPPLSYNDINNALNELNSFGDLINPFDTYKPSKNITTDNERKKFIMEI 456
Db 340 LTOENG-----TLKEINASLNOEKW--NLIQ-----KSESPANYID--EREKSISEL 382
QY 457 KEKIKIEK-----KRIESDKSKSYEDRS---KSLNDITKEYEKILNLIYOSKFPNNIDLINF 509
Db 383 SDQYKOEKLIILROCEETGNAVEDLSOKYKKAQENSKLCELICTSLCEKRNKELDOL 442
QY 510 EKMGMKRSYKVEKLTJHNTFASYENSK--HNLEKLTALK--YMEDYSILRNIVYKEKELKY 566
Db 443 KEAFKKEHQEFLTKL-----AFAEERNQNLMELETVQALREKMDNONSSEAGGLKO 498
QY 567 YKNLSKIENIEITLVENIKKDEOEOLFEEKKTKDE-----NKDEKILEVSDI-- 614
Db 499 EIMTLKEQNKQKQKQVNDLQENEOIMKVMKTRKHCQMLESEPIRINSYKERESEBNOCNF 558
QY 615 ---VKVOYOKVLLMNKIDELIKKTOLILKNVLEK-----HNHIV----- 649
Db 559 KPOMLEVAKEISLDSYVMAOLVQLEAMLNKKELLOESEKERECLQHEIQTIRGDLSTNL 618
QY 650 -----PNSYKQENKQEPYLLIVLKKEIDKLKVMKPVESL 684
Db 619 QDMQSOEISGLKDCILDADEKYSISGPHELSTSQNDNAHLQCSLQVTMKLNE--LEKICEI 677
QY 685 INEEKKNITKESQSPNSPSTBGEITGOATTKPGQOAG-----SALEGDSVQ 731
Db 678 LQAEKYEVLVTEILNDSRSECIY-----ATRKMAEEVGLKLINEVYLILNDDSGLLHGELVE 730
QY 732 ---AQAOEQQAQPPVVPVPEAKAQVTPPAPVNNKNTENVSKDYLEKYEFLNTSYI 787
Db 731 DIPGGEFQEPQNEQHPVSL-----APLDESNS-----YEHLLLS-- 764
QY 788 CHKYLIVSHSTMNKIL--KOYKITEESKLS-----CDPLDLLEFNION 831
Db 785 -DKVEQMHFAELQEKFFLSQSEHKTLHDQHQOMSSKMSLELQTVYDSLKAENLVJSTNLRN 823
QY 832 -----NIPVWYSMFDSLANSLSQLEPMETIEKEMVCNLKYLKNDKYNL 875
Db 824 FQGLDVLKEMQGLGELGVLPSLSSSCVDPSSSLSSIGDSSFYRAL--LEQIGDMSLSLNL 880

```

```

OY 876 LEBKKTSTVKTLLSSSMOPLILTPQODKPEVSA---NDOTSHSTLNLSL-KLEPENTLS 933
Db 881 EGVASNAQCSVDVEFGCSLOEBENTLTKETPSPAKAVELEBLCVEYKQSLKELEBKES 940
OY 932 LG--KKNKIYOELIGOKSSSENYEKTILKSDPFYNESTFNPKSKADDINSINDESKRK 989
Db 941 OGIMKKNKEI-OELEQLLSSEROELDLRKOYUSENOM-----OQKLTSTVLEMSKL 992
OY 990 LEEBINLKKTLTLOSPLYUKYKTLKLEBFDKKKTYGKXKMKQIKKLTLLTBQLESKLSL 1049
Db 993 AAER-----KOTEBLSLEL-EVARLOLOGL-----DLSRSLGIDTEPAIOGR 10355
OY 1050 NNPKHVLQNSVEFENKKEAEIAETENTLENTKILLKH-----1087
Db 1036 N-----ESCDISK-EHTSETTERPKHDVHOICDDAQQOULNDIEKIT 1078
OY 1088 YKGLVK---YNGGSSP-----LKTISE-----E 1108
Db 1079 ETGMLKPTGECSSQSDPTNYERPGECDKTQGSSECSITELSSFGPNALVPMDFLQMOEDIH 1138
OY 1109 SIOTEDNYASLENEKVLKSLGEG--KLKDNLANLEKKKLSYLSGLH---HLIALEKVI 1161
Db 1139 NLQLRVKTENENMLRLHVLIEDDRKVESILN---EMKELDSKILHGEVQMTKIEACI 1194
OY 1162 KKKYVTGNSSENNTPVNNALSEY-----1185
Db 1195 ELEKIVGELKE-NSDISEKLEVFSCDHOELLQVETSEGLNSDLEMHADKSSREDIGN 1253
OY 1186 -----KKFLEPTGDVATVVSSEGSPTLEOSQPKPASTHVAESNTITTSQ--NV 1233
Db 1254 VAKVNDGSMKREPLDVENELSRIRSEKASTIEH-----ALYLEADLEVQVTEKLCLEK 1305
OY 1234 DDEVDVITVPIGESESEDYDIDGOVYT-----GEAVTPS-----VIDNIIISKIE--- 1278
Db 1306 DNEKKOKVIVCL-----EEL-----SVATSERNOLRGEIDLTMSKTTALDOLSEKMEKETO 1357
OY 1279 -----NEKEVLYLPLAGVYSLKOLENNMFPNVAVKILNSRPNKRENFVNVESDLI 1334
Db 1358 ELESJOSECLHIOVVAEAEVKEKTEL---LOTLSSVSELKDKTHLGEKLOS-LEKD- 1411
OY 1335 PYKDLTSSNYVVKDPYKFLNKKER-----DKFLSSYNIKSDITDINFANDVLGY 1385
Db 1412 -SQALSTLTKCELENOQLKMLKREKELLVKESELOARLSSDYEKLNVSKALEALVERGE 1470
OY 1386 YKI-LS-----EKYSDDLS-----IKKTIINDKGENEYKLPFLANNIETL 1424
Db 1471 FALRLSSTOGEVHOLRGRGIEKLVRLAEAEKQOLHIAEKLKEREENDSLKDKVNLERE 1530
OY 1425 YKTYNDRKIDLFVHLEAKVNTYFEKSANE---VKIKEL-----NYLTKTIDOK 1469
Db 1531 LQNSENOQELVIIDAENSKAEVETLKTQIEEMARSLKAVELDOLVTLRSKEKULMTQIOEK 1590
OY 1470 LADPKKNNNEVGLADISTDYNNHNNLLTKFLSTGMVFENILAKTVLSNLDGNIQGLMNLISQ 1529
Db 1591 OGOLSE-----LDKLILSS-----FKSLLEKKQDAEIOIKIE 1620
OY 1530 HQCVKQKOPNSGCFRHLDRRECKCLLVNKGDCVCENP--NPICNENN 1578
Db 1621 ESKTAVEMLON-----OLKELNEAVVALGDDPEIMKATOSLDPPLEEBH 1665

RESULT 13
US-09-310-187A-1
: Sequence 1, Application US/09310187A
: Patent No. 6358751
: GENERAL INFORMATION:
: APPLICANT: Benichou, Gilles
: APPLICANT: Fedoseyeva, Eugenia
: TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
: TITLE OF INVENTION: Graft Rejection
: FILE REFERENCE: UCSF-090
: CURRENT APPLICATION NUMBER: US/09/310,187A
: CURRENT FILING DATE: 1999-05-12

```

;	NUMBER OF SEQ ID NOS:	3
;	SOFTWARE: FastSeq for Windows Version 4.0	
;	SEQ ID NO 1	
;	LENGTH: 1939	
;	TYPE: PRT	
;	ORGANISM: Homo sapiens	
;	US-09-310-187A-1	
	Query Match	4.1%; Score 343; DB 4; Length 1939;
	Best Local Similarity	19.7%; Pred. No. 3,4e-10;
	Matches 301; Conservative	267; Mismatches 563; Indels 396; Gaps 69;
QY	28 LVKLEALEDAVLGTSLFQEKRYNLNCGTAVATTSPTSGKSVASGSGVASGGS 87	
Db	595 LEKNMDPLNETIVV---ALYOKSSLKL-----MATLFs-----SYAADTGDSGSKSGG 639	
QY	88 VASGGSVASGGSVASGSGS-----NSRFTNPSSDSDSAKVADLKHHVPNTLTIKELK 143	
Db	640 KKKGGSFOTVALHRENINLMTLRTHP-----HVR--CIIPNERK 681	
QY	144 YPOLFD--LTNNMLTLCDINGFPKYLDGYEEINELLYK-----LN-----FYED 186	
Db	682 APGVAMDPLVMHQRLCNGVLEIGIRICRGKGP--NRILGDGRQRYRIINPAIPEGDFID 739	
QY	187 LLRA--KLNDVCANDYOQIFP--NLKIRANDLVLYK-----L 220	
Db	740 SRKGTEKLSSLDIDHNQYKFGRHKVFKAAGLLGLEMRDRERLSRIITRMQAQRGLM 799	
QY	221 VEGYRKPLD-----NIKDYK-----KMEDIYK--KMKKTIENTINELIE 257	
Db	800 RIEFKKIYERRDALVIQWNIRAFVGYNWPMMKLYFKIKRPLKSGETEKEKATWKDEEG 859	
QY	258 ESKKTIDKNKATYEFEKK--KLYOAOYDLSTYKQLEAHNLISYLEKRIDLTKNE-- 313	
Db	860 RIKETLEKSEARRRELEKRWLSLQEKMDILO--QVOAEDONLMDAEBCOLLKAKIQ 916	
QY	314 ---NIKELDLINETIKPNPPANGTPNTLDDKNKIEHEKEIKEIKTIKFINIDSFLT 370	
Db	917 LEAKYKENNELDEBE--MNA-----ELIAKKKKLED--ECSELKKDI----- 956	
QY	371 DPLELEYLRBKNNKIDISAAY--TKESTEPNEYNGVTYPLSYINDINNALNELSPGLD 429	
Db	957 DDLELTL-----AKVEKKNATE-----NKVNLTPEEMAGLDEI 990	
QY	430 INPFYTPEPSKNITYTDNERKKFFINEIEKIKIEKKKLESOKKSVEDSKSLNDITKYE 489	
Db	991 TAKL-TKE-----KKAOLEAQIQ--ALDDJOVEKD-----VNSLSKCV 1027	
QY	490 KLNLMEIYSKFNNNIDLTNFEEKMMKRRYSUYE--KLTHNHTFASYENSKHNLEKTRKA 546	
Db	1028 KLEGQVVD--LEGSLQEGKKVAMDERAKRKLEQDKLT-QESIMDLNDKLOIEPKLKK 1084	
QY	547 LKY-----MEDYSLNIIYVERKLTAKYYKLNLSIKTENMEITLVANIKKDEQLPEKIT 598	
Db	1085 KEFDINOONSKIIEEQALALOLOKKLKKNQARIEELEBELE--AERTARAKVKELRSOLS 1142	
QY	599 KDEKNPDKEILEVSDIYAIVOXYVLNMK--IDELKTKQILINKVELNHNIHPASYKOE 656	
Db	1143 RELFEISRLEBAGATSVOIE----MKKKRAEFQCKMRRLDEEFLQHEA-----TAAL 1194	
QY	657 NKQEBEYLIVALKETDKLVMPKVESLINEEKKNIKTGESODNSEPSTEGETIGQATTK 716	
Db	1195 RKKHADSVAELGEQIDNIO-----RVKQKLEKEKSEFKLE--LDVTSNMEOITIKAKMIE 1248	
QY	717 PGQAGSALBEDGSVOAQAOEQOKAOPRPVVPYPRKAAV-----PPRPAVVNKTENVSKLD 773	
Db	1249 KVSR-----TLEDAQNER-----VKLEEQRSINDFTTORAKL-QTENGELAR 1291	
QY	774 YLEKUYEPLNTSYICHKLIYVSHSTMNE-----KILAQYKIREESKSLSS----- 819	
Db	1292 QLEE-----KALLISQLTRGLSLTYQQOMEDKLROLEEBGKAKMLAHLAGOSAR 1339	
QY	820 --CDPLDLLENIONNIPVYSMFDSLNSLSQLEMEIYEKEVCNLYLKLD--NDKIKNLL 876	


```

Db 1228 KL--QTENGELARQLEEKALIMQITRGKL-----SYTQOMEDLKROL---EEEGKA 1274
Oy 818 SS-----CDPLDLLEFNIONNIPVYMSFMDLNNLSQLEPMETYEKEMVCLNK 865
Db 1275 KNAHLALASARHDCOLLLEQYEEEMAKAELOVRYSKANSFVAQMRTRYETDAIORTEE 1334
Oy 866 LKD-NDKIKNLEAKKRVSTSVKTLSSSSMOPSLTPQCKPE--VSANDTSHSTNLNN 921
Db 1335 LEEAKKLAQRLODAEEAVEAVNA-KCSSLEKTKHRLQNEIEDLAWDVERSNAAALAK 1393
Oy 922 SLKLFENILSLGKNKIYQELLGOKSS-----ENFEKILKDDSTYYNNS 966
Db 1394 KORNEFKITAEWKQK--YEESOSELESQKEARSLSTELFKLNAYESLEHLETKREN 1451
Oy 967 FTFNVKSKADINSLND-----ESKRKLEEDINKLKTL-----1001
Db 1452 -----KNLOEELISDLTEQJGEGGKNVHELEKIRKOLEVEKLELOSLAEASLEHEBECK 1506
Oy 1002 ----QLSPDLNKKYKLEER-LFDKKKTVGKYKM-OIKKLTLEKOELESKLSLNNPKHV 1055
Db 1507 ILRAQLEF---NQIKAEIERKLAEKDEMEMOAKRNHLRVVDSLQTSILDATRNRNALRY 1563
Oy 1056 LQNFSEFEKKKEAEIAEENTLENTKILL---KHKKGVKKYNNESSPKLTSESTQ 1111
Db 1564 -----KKMEGDLNEMEIOLSOANRIASEAQKHLK-----NAQAH---LKDTOLQ 1605
Oy 1112 TEDNVASLENFKVLSKLECKLDNLNLEKKKLSYLSGGLHILAELEVIKKNKYTGNSP 1171
Db 1606 LDDAVANAND-----LKENAIVERNRLLOALELEELRAVVEQERSKLAQEL 1655
Oy 1172 SENNDV-----NNALESYKFL-PEGTDVATVSESSDTELOSOPKKPASTHWGAE 1224
Db 1656 IETSEVQOLHSQNTSLINQKKKMDADLSQLOTEVEEAQOECHNAEEKAKKATIDAMMA 1715
Oy 1225 NITTSQ-----NVDEVDVITVIRFGESEBDYDQGV-VIGEATPVSVIDN 1272
Db 1716 EELKEQDTSALHERKKKMEQTIKDL-----QHRLDEAQDLALGGKKOLOKLE 1766
Oy 1273 ILSKIENEVEVLYLKPLAGYVRSKLEQENNVMTFVNVYKDIINSFKNREKNVLESD 1332
Db 1767 RYRELENELE-----AEQKRNASVYGMRKSEKRIKELYQTEEDKKN-----1809
Oy 1333 LIPYKDLTSSNVYKDPYKFLNKKRDKF---LSSYNYIKDSTID 1373
Db 1810 LVRLQDLVLDKLOLKVAKYKROAEAEQANTNLSKFRKYQHEID 1853

RESULT 15
US-09-104-324B-4
; Sequence 4, Application US/09104324B
; Patent No. 6232460
; GENERAL INFORMATION:
; APPLICANT: T rect, Ozlem; Sahin, Ugur; Pfreundschuh, Michael
; TITLE OF INVENTION: Methods for diagnosis and treating cancers,
; TITLE OF INVENTION: And methods for identifying pathogenic markers in a sample of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,324B
; FILING DATE: 25-June-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 08/892,702
; FILING DATE: 15-July-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6232460man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3000
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-09-104-324B-4

Query Match 3.9% Score 332; DB 4; Length 976;
Best Local Similarity 20.7%; Pred. No. 5.6e-10;
Matches 236; Conservative 204; Mismatches 354; Indels 344; Gaps 59;

Oy 196 CANDYQIIPF---NLKIRANEID--VLKKLVGYRRPLDINKDNYGKMEYIKKRNK-- 247
Db 41 CTFEDLEFPRAKNLSKNGENIDSDPALQKVN---LP---VLEQYGNDSCHYQELKOS 94
Oy 248 TIEN-----INELLESKTIDKKNKATKEE-----EKKLYQAQ---YDLSI 287
Db 95 DIENSEGLSRVFSKLYKKEEK-IKKKVSTFELRLROKESKLOENRKIIIEORKAIOELQF 153
Oy 288 YNK-----QLEAHNLISVLEKRIQDLTKKNENIKELDKINEIKNPPRANGNPTNLLQK 343
Db 154 GNEKVSLEKEEG-----IQENKDLIKENNAIHLNCLNLE-----TCARSAEK 196
Oy 344 NKRIEHEKEIKELIANTIFENIDSLFT-----DPLELEYLREKKNKI-----D 387
Db 197 TKKYVERETRTQVYNDNNINIEKMTTARQELAVQANSLSLEHNRFLKDEYKELQHLQED 256
Oy 388 ISAKVETKES-----TERNEYNGVYTPYLSYNDINNALNELNSFGDLINPRTYTE 438
Db 257 YKKEINDKEQVSLDILQITEKENKKMDLFTLL--ESRDKVQMLE-----EKTKL 305
Oy 439 PSKNITDNERKKFIKEIKIKIEKKKIESDKSYED---RSKSLNDITREKELNVE 494
Db 306 QSENLQOSTEKQHLTKLELDIVISLQRSYSTOKALEBDIQATKYTIQOLTEKETQME 365
Oy 495 IYDSKFNNDLTNFE-----KMGKRSYKVEKTLHNHTFASYSNKNH--LEKILTRA 546
Db 366 SNKARAHSFVYVEFTYQSLSEELLRTQQRLEKKNEDQKILTMELQKSSLEEMTKL 425
Oy 547 L--KYMEDYSLRNIVVEKELKYKNLISKIENIEFTLVENIKADEQLPEKKTIDENKP 604
Db 426 TNKKEVELEKELKVLGEKFTLYEN-----KQEKIAEELKGTQEDQLGLQAREKEVN 479
Oy 605 DEKILEVSDIVKVOQVYLLMNKIDELKLTQL---ILKNVELKHNHIVPNSYKOEKQOR 661
Db 480 D---LEI-QLTAITTSEQYISKVEVKD-KTELENEKLTETLSHC---NKLSENKE-- 529
Oy 662 YLLIVLKEIKDLKVPMPYKESLINBEKKNIKTEGOSDNPSTSEGETGQATTKPGQQA 721
Db 530 -----LTQETSQMTLELKNQOEDINNKKQOEEMLKQIENLQIETEQLRNE----- 575
Oy 722 GSALBEDSVQAQOEQKQAPVPVPRPEAKAOPVPPAPVNNKTENV--SKLDYLEKLYE 780
Db 576 -----LEYVREELKQK-----DEYKCKLDSSECNMLRKQOVENKKNYIEELQ 620
Oy 781 FLNTSYICHHYILVSHSTNNKILKOYKIKTEESKLSQCDPLDLLEFNIONNIPVYSMF 840
Db 621 -----ENKALKKKGTAEKQNLNVEI---KVNKLE---LEL-----650
Oy 841 DSLNNSLSQLFMEI---YEKEWVCNLKYLKDNKI--KNLLEAKKRVSTSVKTLSSSSMQ 895
Db 651 -----ESAKQKFGBITDYOKEI-----EDKKISEENLEBEVK---AKYIAQAV-- 693
Oy 896 PLSLTPODKREVSANDTSHSTNLNNSLKFENILSLGKNKIYQELLGOKSSSENFYEKI 955

```

Db 694 -----KLOGEIDKRCOH-----KIEMVALMEKHKHQOYKIIERDSE----- 731
Qy 956 LKDSOTFTNESTNTNVKSKADDINSLNDESKRKLEEDINKKTKLQLOSPDYKNYKTKL 1015
Db 732 -----LGLYKKEOEOSL-----RASLEIEISNLKABL-----LSVKOLEI 769
Qy 1016 ERLPDKKTVGKYKMOIKKLT-LKEOLESKLSLNPNKHYLQNSVFENKKKEAIAET 1074
Db 770 ER--BEKE--KLKKEAKENTATLEKKDK-----TQTF----- 799
Qy 1075 ENPLENTKILKHXYKGLVYKYNSSPLKTLSEESIQTEDNYASLNEFKVLSKLEGKLD 1134
Db 800 --LLETPETIWK-----LDSKAVPSQTVSR-----NFTSYDH----- 829
Qy 1135 NLNLEKKKLSYL-SSGLHLHLAEKVIKKNKYTGNSPS-----ENNTDVNNALSEYK-- 1187
Db 830 --GISKDRDYLMTSAKKTLSTPLP-----KAYTKLPTPKLQORENLNPIRESKKR 882
Qy 1188 -----FLPECTDVATVSESGSDTLBO-SQPKKPASTHVGAE-----NTITT 1229
Db 883 KMAPEFDINDSSSETTDLMSVSE--ETLTKLYNNPNPASHLCKVTPKKAAPSULT 938

RESULT 16
US-09-134-001C-3242
; Sequence 3242, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NOCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3242
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3242

Query Match 3.78; Score 314.5; DB 4; Length 1151;
Best Local Similarity 19.78; Pred. No. 5.8e-09;
Matches 268; Conservative 225; Mismatches 485; Indels 379; Gaps 63;

Qy 252 INELIE--ESKRTIDKKNNA---TKREEKKKLYAOYDLSIYKQOLEAHNLISYLEK 304
Db 9 INEQDITLYLKKLEKESNNRPLINTINOVIKKYLLNOYTATFVGHSFGSKTLLINLLE 68
Qy 305 RIDTLKKNENIKELDKINEIKNPPANSNGMTPTLLDKNKI-----EEHEKEIKELAK 359
Db 69 Q-DILP-----SSPVPTS-NTAIVSAKEDIIANLQOQYTKLKTND 111
Qy 360 TIKFNIDSLFDPLELEYLEKKNNT-----DISAKVEF--KESTERPEY--PAGVTPPL 411
Db 112 VKOMRQWVDEESIEINPSNKFNGFTFODTPGVDSVNAVATQOSTEOMYTSNLLFTYV 171
Qy 412 SYNDNNALN-----ELNSFGDLINRPDYTKPEPSKNIYTDNRKKKFINIEKIKIE--K 464
Db 172 DYNHOSALNFKMKRINEVGIP-----IFVINOIDH-NE--EELTFETPK 216
Qy 465 KRIESDKSYEDRSKSLNDITREYKELLINEIYDSKFNNNIDLTNEKMKMGKRYSYKVEKL 524
Db 217 SRVEKSIDW-----DI-----KLQDYVYSKF----- 239
Qy 525 THHNTFASYENSKHLEKLTAKLTM-----EDYSLRNIYVEKELKYKMLIKIENE 577
Db 240 -----DHPQNDIKLSNFLVMDQHRSTEDYVNRFTIOFIYDAQYI-----YONE 285

Qy 578 IETLVENIKKDEEOLPEKKITKDEKNPDEKILLEVSDIYKVQYKYLMMKIDELKKTOLI 637
Db 286 MOSTIDTLOINEEOPREAVIOPQNO-----EVSABOILNDSNOQFNYLQOKR----- 335
Qy 638 LKNVELKHNHVPNSYKQENKOEPPYLLIVLKKIEIKLKLFMPKVESLLINEEKNKIKTEGO 697
Db 336 ----DILDNAVIMTYDMRESLNR--YLESMAVDF-----KVGQFNRKKK----- 374
Qy 698 SENSEPSTEGEITGQATTKPGQOAGSALEGDS-----VQAOAQEOQAOPVPVPVPEBA 751
Db 375 KEEQDIKRLNEATYLOEKNVNOQVROPLEDEMSFTLRFINKHAVNEK----- 421
Qy 752 KAQVTPRAPVNNKTEWNSKLDYLEKIEFLNTSYICHYILIVSSTWNEKILKQYKITK 811
Db 422 -----ILQEDVYPSL--ISELVO--TQTSISNTYVL-----TSDEVIKALNKKI 464
Qy 812 EESKLSGCDPLDLFNIONNIPVWYSMEDSLNNSLSQLFMEIYERKENCYLKLDNDK 871
Db 465 EMES-----TPLEAVNHHQV-----NELSSDENEDRYEYDRYTELNTLKDSL 509
Qy 872 IKNLLEAKKYSTSVKTLSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLEPNILS 931
Db 510 SHNYKHYIHLDDSLDKLIGRTETHEFLQENSTAYHRKHEQOHRNEFTYSNQDIKRALD 569
Qy 932 LGKNKNIYOELIGOKSENFEYKILK-DSDTFYNESFTNFVSKADDINSLNDESKRKL 990
Db 570 IYKDVPLF-----DRTKODITITILRDNOITKVGFGTFSGKSSLINALLGDN---YL 621
Qy 991 EEDIN-KLKKTLQLSFDLYNKKYKLELRLFDKKTKVGYKMOIKKLTLLKEOLESKLSL 1049
Db 622 VSPNPTAATATLSEKESQITLK-----SKEQILLEVN-- 656
Qy 1050 NNPKVYLQNSVYFPKKK---EAEIAETNTLENTKI-----LKHXYKGLVYKYN--ES 1099
Db 657 ---HVLEFEYLSFTLQNFISDKLEKLSKLEKQOLAFISAIEKH--EMYSMLEH 708
Qy 1100 SPLKTLSEESIQ---TEDVYASLENFYVLSKLEGKLDNLNLEKKKLSYSSGLH----- 1151
Db 709 SLIHVVSLEIKKMSAEDYATF--VKTY-----HLKPLDMLKKKIILDISGLHSNNQR 761
Qy 1152 ----HLIAELKVIKKNKYTGNSPSENNTDVNNALSEYKFLPBGTDVATVYSESGSDT 1206
Db 762 HTNETEQILTSDDLILVYVYFNHSFTDND-----KAFIEHMKD-----NO 802
Qy 1207 LEQSQPKKPASTHVGAESEITITTSQNVDEVDVYIIVPFGSEEDY--DOLGOV----- 1259
Db 803 INENQAFK-----MIINAVDLAEDKODI-----QAVEDYADALGOVNLHSD 844
Qy 1260 VGEAVTPSVIDNLSKIENEYEVLYKLPLAG---VYRSILKQLENNVATFVNVNKDI-- 1314
Db 845 IYSVSSROSLSNKNNGINELRESIOYFAKVESRTILDEQMTYQLOOMNNSFFKMKIKDFPD 904
Qy 1315 ---LNSRPKNREFNK--VLESDLIPYKDLTSSNVVAKPKPLNKKERKDFLSSYNYI 1368
Db 905 DNAKLSARQKLNHYKNOTLILNOELIDTQRTFNEVEBOYV-HLNERLKLOLLE---V 960
Qy 1369 KSIDTIDINFADVILGYKILSEKKYSOL-----DSIKKYINDKGENEKYLP 1416
Db 961 KSVFNSQMTQNNDFNEKKISPKIYLDQHORLLEQSLITERIKKYFNSQL--BEQILP 1018
Qy 1417 FLNNIETLYKQVNDK-----IDLEFVHLE-----AKVNIYYEKSNEYV 1455
Db 1019 VAKKLNQIHVITINAKFNVNPSLVYDALLOIELNSMLQSLPKOLTKRKIYN---PKSQDI 1075
Qy 1456 KIKELNLY-LKTIQDKLADFKKNNNEVGIADLSTDYNNH 1491
Db 1076 QVHIANQTELLLODDLNSLRQLN-----DYIYH 1103

RESULT 17
US-09-134-001C-4820
; Sequence 4820, Application US/09134001C


```

: REFERENCE/DOCKET NUMBER: 4600-0111.30
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 324-0880
: TELEFAX: (415) 324-0960
: INFORMATION FOR SEQ ID NO: 51:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1312 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NF.
: INDIVIDUAL ISOLATE: 389 TO 4324
: US-08-687-080-51

```

Query Match 3.7%; Score 308; DB 2; Length 1312;

Best Local Similarity 19.4%; Pred. No. 1.5e-08;

Matches 248; Conservative 214; Mismatches 453; Indels 364; Gaps 56;

```

Qy 411 LSYVDINNALNELSFGDLI--NPFDYTKRPSKNITDNRKRFINKIKIKIEKKKI 467
Db 84 LQFQDVN-----GELAVQSRSMVCTQSKRTEFKTGTTRTHGKVSLSKCA 134
Qy 468 ESDKRYEDRSK--LNDITREYE-----KILNEYDSKFNNDLITNEKMMG 514
Db 135 EIDEMISSSLGSKAVLNNTVFCHEODSNMPLSEKALKOKFDELF--A 182
Qy 515 KRYSYVEKLTHNTFASYSKHNLEKLTALYMEDY--SLRNIVYEK--LKY 567
Db 183 TRYIALETTLROVQOQOKVEYQME--LKYLOKFKACEIRQIISKEAQLSS 237
Qy 568 KNLISKIENETLENTVENIKKBEQOLFEEKITKDEKPRKILEVSDIYKVOK--V 622
Db 238 KEIVKSYENELDPLKNRKLEIENL--SKIMKLNE--IKALDSRKKOMKDNSELE 290
Qy 623 LLMKKIDELKKTQILKKNVLEKHNHIVNSYKOEKOEPPYLIYKREIDKLYMPKVE 682
Db 291 EKMEKVEQDGEOL--NDLYHN--HQRVREKERK--LVQCHRELEKLN--KES 336
Qy 683 SLINEEKNKITEG--OSDNE--PSTGEITGQ 712
Db 337 RLNOEKSELLVEQGRLOQADRHQENHIRADSLIQSLATQLELDFGRGPFSEQIKNF 396
Qy 713 ATTPGQAGSA--LEGDSVQAQAEQKQADPPVPPVPEAKQVPTTPAPYNNKTE 767
Db 397 HKLVREOGEAKTANOLMDFAKETLKQKQ--IDEIRDKKTGLGRITIELKSE 448
Qy 768 NVSKLDVLEKLYEFLNTSYICHYILVSHSTWNEKILKQYKITEEESKLSSCDPLDLF 827
Db 449 ILSK-----KONELKNV-----KTELQOLESSDRILELODKAEKELSKAE-- 493
Qy 828 NIQNNIPVMSMFDLSNLSLSQLEMEIYEKEMVCNLYKLKDNK--IKNLEEAKK 881
Db 494 -----NSNVETLKMEVLSIQ-----NEKADLDRTLRKLDQEMEQ 527
Qy 882 VSTSVKTLSSSQMPLSTPDQREVSANDTSHSTNINSLKLFENILSL--GKNK 936
Db 528 LNH--HTTTRTQMELTKRKADKDEQIRKIKSRHSDELTSILGYFPNKKOLEDMWLSKSK 585
Qy 937 NIYO-----ELIGOKSENFYEKILKSDPTFYNESTNPFVSKADINSLSNDSK 986
Db 586 EINTORRLAKLKEKLASSSEONKHNINELKRKE--EQLSTVEDKLFVDCGSDOFFSD 641
Qy 987 RKLLEEDINKLK--TLQISFDLYNKKYKL-----ERLFDKKTGVGKYMQI 1032
Db 642 LDRLEKEIEKSSKORAMLAGATAVYSQFITQLTDENOSCCPVQCRVQTEMLQEVISDL 701
Qy 1033 K-KITLKEQLESKLSNINANKH-----VLQNFSVFFPKKKAEALAELENTLENTKI 1083
Db 702 QSKRLAPDKLTSESELKKEKRREDEMLGLVPRKOSIIT--DLKEKEIPELRNKLQNVNR 759

```

```

Qy 1084 LKHKKGLVKYYNGESSPLKLTSEESIQTEDNVASLENFVYLSKLEGKLDNMLEKKL 1143
Db 760 DIQRLKNDIE--DEETILGITIMPE--ESAKVCLTPTVITMERQOMELKQ--VEKRI 809
Qy 1144 SYLSSGLHHLJAELEKVEIKRNKNTGNSPSENNTDVNNALESYKFLPGTGVATVVSSEG 1203
Db 810 AQQAALQGI--DIDRTVQVQVNOEKQKQHKLDIVSSKIELNRKLI----- 853
Qy 1204 SDTLQSQPKKRPASTHVCAESNTITTS--QNVDEVDVYIYVIFGESEEDYDLDQV 1259
Db 854 QDOQOIQHLEKSTNELSEKLOISTNLRQOLEEOT--VELSTEVQSLYREIND- 907
Qy 1260 VTGEAVTPSIDNILSKIEENEVLYLKLPLAGVYSLKOLENNWTFPVNWKDILNSRF 1319
Db 908 -AKEVSP--LETLEKEQOE-----KEELINKNTSNKTAQDKLNIRK 948
Qy 1320 NKRENFKNVLESDDLIPYKDLTSSNYYV--KDPK-----PLNK-----EKRDPLS 1363
Db 949 EKVKNIHGYM-----KDI--ENVIQDKDYKKQKETEINKVYLAQLSECKHREKINE 999
Qy 1364 SYNYIKDSIIDP-----INFANDVL-----GYKILS-----EK 1392
Db 1000 DMRLMRQDIDTQKIQERMLQONLTLRKRNBELEKVEEERKQHLKEMGOMVLOMKSEHQ 1059
Qy 1393 YKSDLSIKKYYINDKGENEKY-----LPLNIIETLYK-----TYND 1430
Db 1060 LEENDINIKRNHNLALGQKQYEEBIIHFKKELRPPQRRDAEKYREMMIYRTTELYNK 1119
Qy 1431 KIDLFVHLEAKVNLVYTEKSNVEYKIKELNYLKTIOQKLADFPKNNNF----- 1479
Db 1120 DLDIYYKTLDOAIMKFH-----SKMKEIN--KIIRDLMSRYGQOIEYIEIKSDAE 1171
Qy 1480 -VGIAADLSTDVYNNHNLTK 1497
Db 1172 NVASDKRRNRYRVVMLK 1190

```

RESULT 19

US-08-592-126-148

Sequence 148, Application US/08592126

Patent No. 5821091

GENERAL INFORMATION:

APPLICANT: Gregory Doljanov

TITLE OF INVENTION: Transcripts Encoding Immunomodulatory

NUMBER OF SEQUENCES: 151

CORRESPONDENCE ADDRESS:

ADDRESS: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/592.126

FILING DATE:

CLASSIFICATION: 435

AUTHOR/AGENT INFORMATION:

NAME: Shultz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 148:

SEQUENCE CHARACTERISTICS:

LENGTH: 1312 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
US-08-592-126-148

Query Match 3.6%; Score 307; DB 2; Length 1312;
Best Local Similarity 19.4%; Pred. No. 1.7e-08;
Matches 248; Conservative 214; Mismatches 453; Indels 364; Gaps 56;

QY 411 LSYNDINNALNELNSFGDII--NPDYTKPSKNITYTDERKKFTNEIKIKIKKKI 467
DB 1 LSYNDINNALNELNSFGDII--NPDYTKPSKNITYTDERKKFTNEIKIKIKKKI 467
DB 84 LQFRDYN-----GELIAVQRSMVCTQKSKTEFTLEGTITRTKHEKVSLSKCA 134
QY 468 ESDKSYEDRSK--LNDITKEYE-----KLMEIYDSKRNINIDLTNPEKMG 514
DB 135 EIDREMISSLGVSKAVLNNVIFCHQEDSNMPLSEGRALKOKDEIRFS-----A 182
QY 515 KRYSYVEKLTHTNPFASYENSKHNLEKTLKALYMEDY-----SLRNIYEKE--LKY 567
DB 183 TRYIKALLETROVROGOKVKEYOME-----LKYIKYKERKACETIRDQITSEADLTSS 237
QY 568 KNLISKIENEIETLVENIKKDEBQLEFKKITDENKPRDEKILEVSDIVAVOYOK-----V 622
DB 238 KEIYKSEYENELDPLKNRLEIEHNL--SKIMKLNE-----IKALDSRRKKOMEKDNSELE 290
QY 623 LLMNKLDELKQQLILKANVELKHNHVPSYKQENQEPYLLILVLEKEIDKLKVFMPKYE 682
DB 291 EKMEKVFQSTDRQL--NDLYHN--HORTVREKERK-----LVDCHELEKLA--KES 336
QY 683 SLINEKKKNIKTEG-----QSDNSE-----PSTEGEITGQ 712
DB 337 RLINQKSELVLEQGRLOQADRHQEHIRARDSLQSLATQLELDGFERGPFSEGRQIKNF 396
QY 713 ATTKPGQOGSA-----LEGDSYQAQOEOKAQAPVPVPRPAKQVPRPAPVANNKTE 767
DB 397 HKLVBRQEGEAKTANOLNDFAEKETLKOQ-----IDIRPKKTKGLGRHIEKSE 448
QY 768 NYSKLDYLEKLYEFLNTSYCHYLLVSHSTMEKILKQYKITEESKLSGSDPDLLE 827
DB 449 ILSK-----KQNELKAV-----KYELQOLEGSSDRILEDDQELIKARELSKAK----- 493
QY 828 NIONNIPVWYMSFDSLNSLSQLFMEIYEKWCNLYKLKDNK-----IKNMLEAKK 881
DB 494 -----NSNVELTKMEVISIQ-----NEKADLDRLRLKLDQEMEG 527
QY 882 VSTSVKTLSSSSKQPLSLTPQDKPEVSANDDTSHSTNLSKLEFNILSL-----GKNK 936
DB 528 LNH--HTTTRTQEMELTKKADKDEQIRIKSRHSDELTSILGYPFNKKQLEDMLHSKSR 585
QY 937 NTYQ-----ELIGQSSSENFEEKILKDSDTFYNSFTFVFSKADINDLSDESK 966
DB 586 EINQOTDRILAKNLKELASSEQKNHNNINNELKRE-----EQLSSEYKLDVYCSDQPESD 641
QY 987 RKLLEDINKLKK--TLOLSPDLNKKYKLT-----ERLFDKKTKVGYKQMOI 1032
DB 642 LRLKREIKSSKQRAMLAGATFAYVQSOFITQULTDENQSCPCQVQFQEAELQEVISDL 701
QY 1033 K-KLITLKEQLEBSKLSLNNPKH-----VLQNSVFPPNKKKEKELIAETENTLENTKI 1083
DB 702 QSKRLAPRKLKSTESLEKKERKRDMLGLVPMROSII--DLKEKEIPELRKRLKLVNVA 759
QY 1084 LKHNYGLVKNYNGESSPLKLTSEESIQTEDNVALENKRYVLSKGLKLDKNLNLKKKL 1143
DB 760 DIQRLNDE--EDETLLGTIMPEP--ESAKVCLDTVTIMERFQWELKD-----VERKI 809
QY 1144 SYLSGGLHLILAELEKVINKNKNTGNSPBNNTDVNNALESYKPPDEGDTVATVASEG 1203
DB 810 AQQAALAQGI--DLDRTYQVQVNDQEKQKHKLDTVSSKLELNKLL-----853
QY 1204 SPTLEOSQPKKPASTHVGAESNTITTS-----QNVDEVDVILVPIFGESEEDYDLAGOV 1259

DB 854 QDOQEQIQHLKSTTNELKSEKLEQISTNLOROOLEBOT-----VELSTEQSLYREIKD- 907
QY 1260 VTGEAVTPSVIDNILSKINEYEVLYLKLPLAGYRBLKQLENNVTFVANNKIDLNSRF 1319
DB 908 -AKEQVSP--LETTLEKFOE-----KEELINKKNTSKIAODKINDIK 948
QY 1320 NKRENFKNVLESODLIPKDLTSSNYV--KDPK-----FLNK-----EKRDKPLS 1363
DB 949 EYKKNHNGM-----KQI--ENYIQDGKDDVKQKQETELNKNVLAQLSBCEKREKINE 999
QY 1364 STNYIKDSIDTD-----INFANDVL-----GYTKILS-----EK 1392
DB 1000 DMRIMRODIDTQKIQERWLQDNLTLKRNBELKEVEERKQHLKENGQAOVLOMKSEHK 1059
QY 1393 YKSDLSIKRYINDKQGENKY-----LPLANNIETLYK-----TYND 1430
DB 1060 LLENIDNIRNHNALGROKGYEEELIHKRELREPOFDAEKEKYEMMIVARTTELTVK 1119
QY 1431 KIDLFYIHEAKVLANTYEKSNEVVKIKELNLTQDKLADKKNNF----- 1479
DB 1120 DDIYKTLQALMKFH-----SMKMEIN--KIIRDLWRSTYRGODIEYIEIRSDADE 1171
QY 1480 -VGADLSTDYNNHNLTK 1497
DB 1172 NVSASDKRRNYRYVYMLK 1190

RESULT 20
US-09-541-782-4
Sequence 4, Application US/09541782
Patent No. 6284480

GENERAL INFORMATION:
APPLICANT: Nislow, Corey
APPLICANT: Sakowicz, Roman
APPLICANT: Berard, Christophe
TITLE OF INVENTION: Antifungal Assay
FILE REFERENCE: 1015
CURRENT APPLICATION NUMBER: US/09/541,782
CURRENT FILING DATE: 2000-04-03
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1038
TYPE: PRT
ORGANISM: *Saccharomyces cerevisiae*
US-09-541-782-4

Query Match 3.6%; Score 303.5; DB 4; Length 1038;
Best Local Similarity 21.2%; Pred. No. 2e-08;
Matches 229; Conservative 169; Mismatches 395; Indels 287; Gaps 50;

QY 221 VEGYRRPLNIKQNVG--KMEYIRKNNKTIENINELIESKKTIDKNNKATKEEEKKL 278
DB 132 VEGRGASQDLIPDEVAGPLQDFIKGYNCTV-----LVYGMTST-----GKT 173
QY 279 YQAQYDLSTYNNQLEBAHMLI--SVLEKRIDTLK-----KNENIKELLDKI 322
DB 174 YVTGDEKILYNGELSDAAGIIPRVLLKLPDTLELQONDYVVKCSFIEYNEELKDLDD-- 231
QY 323 NEIKNPPRANSQNTPTLDDKNNKIEHEKEIKELIKTKFN-----DSLFTDP 372
DB 232 -----SNSNOSSTGPD-----QOFMKKLIFASSTANNNTTSASASSSRSHNSSP 278
QY 373 LELE-----YLRKKKNNDISAKVETKESTEBNEVPNGVTYPLSYNDINNA--LN 421
DB 279 KSLNDLTPKAALIRKLRLKSLPTTKQYQOQOAVNSRNNSSN--SGSTNNASSNTN 336
QY 422 ELUSFGDLINPDPYKPSKNITYTDERKKFTNEIKIKIKKKI-----KIESDK-KSYE 475
DB 337 TNNQGSAAHPNDQT-----NGITIQWLOEPHITNMAEGNLQGLKHNQVASTKANDPS 392
QY 476 DRKSLNDITYKEYEKLNEIYDSKFNNNIDLTNFERK-----M 513

```

Db 333 SRSHTIFITTL-YKKHOBELFRISKMLVDLAGSENINSGALNORAKAGSINOSLLTL 451
Oy 514 GKRSYKYEKLTHTNTFASYSKSNHLEKTKALKYMEDYSLRNIYVEKELKYNNLSK 573
Db 452 GRVINALVDKSGH-----IPFRES-----KLTRLQO-----DLSGN 483
Oy 574 IENEIETLVENIKKEBOL-----FEKKTIDENKPKDEKILEVSIYVQVKVLLMKI 628
Db 484 TKTALIAITSPAKVTSEETCSTLEVASAKAKNNKPOLGSFTMKI-----LVKNIT 535
Oy 629 DELKKTQILKVELKHNHIV-PNSYKOENKOEYLLIVLKEIDKLVMPKVESLINE 687
Db 536 MELAKIKSDLLSTKSKESGIYSQDHYKNLNSDLESY-----KNEVOECK---REISLSK 588
Oy 688 EKKNKTEGSDNSEPSTEGEITGOATTKPGQAGSALGDSVOAQOQOAPVPVP 747
Db 589 NALLVKDKLKSKEITQSONCOIESLKTIT-----DHLRQQLDKQKTEIEI--- 634
Oy 748 VPEAKAQPPTPPAPVNNKNTENVSKL-----DYLEKLEFLNTSYICH--KYIIVSHSTM 799
Db 635 -----SDFNKNTLOKLTETVOMALHDYKRRELD-LNOKFEMHITKEIKLKSTL 681
Oy 800 NEKILKOYKTKKEESKLSGCDPLDLFNIONNIPVMSMFDLSLNSLSQLEFMEIYEREM 859
Db 682 FLQI-----NTMOQESILOET-----NIQPNL-----DMIKNEVLTLMRTMQEKAE 722
Oy 860 VCNLYKLDKNDKIKNLEAKKVTSTYKTLSSSMQPLSLTPQDKEPVSA--DQTSHT 917
Db 723 L--MYK-----DCVKILNESPFENVV-----IEKIDILRVDPQKFKYKNAEMLSDIS 769
Oy 918 NLNNSLKLEFENILSGKN---KNIYOELIGOKSSENFYEKILKSDTYNESFTPVSK 974
Db 770 EENNMMKOYL-----KNHFFKNHQBELL-NRHVSTYENIEKRT---NEVEVFKVYL 818
Oy 975 ADDINSLNDESKRRKLEEDIKLKKTQILSFDLYNKYKLELRLFDKKKYVGK-----YK 1029
Db 819 NDHL-----DENKKLIMHMLTTATSAVIDQEMDLFEPRKRVKENSFDLINDCDSMNEFYN 874
Oy 1030 MOIKKLTLEKQLESKLSLNNPKHVLONFVFFNKKKEADIAETENT---LENTKILK 1086
Db 875 SMAATLSQIKSTVDTSSNSM-----ESISVM-----KGOVESENAISLKNNTKFN 923
Oy 1087 HYKGLVYKYNGESSPLK---TLSEESIQTED-----NYASLENFVLSKLEGL 1132
Db 924 QFEDLINKNHMLKDNIKNSITSTSHITNVDDIYNTIENIMKNYGNKKNATKDEMIENIL 983
Oy 1133 KDNILNEKK---KLSYSSGILHLLIAELKEVYKKNNTGNSPSENNTDVNNALESYKFL 1189
Db 984 KEIPNLKMKPLRLSNINSN-----SVQSYISPKHAIDEKNSSENVDN--EGSRKML 1035

```

RESULT 21

```

US-09-723-820-4
; Sequence 4, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1038
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-723-820-4

```

```

Query Match 3.6%; Score 303.5; DB 4; Length 1038;
Best Local Similarity 21.2%; Pred. No. 2e-08;
Matches 229; Conservative 169; Mismatches 399; Indels 287; Gaps 50;

Oy 221 VGYRRKPLDNIKDNV--KMEYIKRKNKTTIENINELIEESKRTIDKKNATKEEKKKL 278
Db 132 VGPASODLIPDEVAGPLFQDFIGYNGCTV-----LVYGMST-----GKT 173
Oy 279 YQAOYDLSTYKQLEAHNLI-SYLEKRDITLK-----KNEIKELDDKI 322
Db 174 YTMTEDEKLYNELSDAGIIPRVLLKFLDLEQONDYVVKCSPELEYNEELKDLDD-- 231
Oy 323 NEIKNPPANSNGPNTLDDKMKLIEEHKEKEIETAKTIFNI-----DSLFTDP 372
Db 232 -----SNSNGSSNTGFD-----GQPMKILRITASSTANTTSNSASSSNSRNSSP 278
Oy 373 LELE-----YLDREKNKNIDISAKVETKESTEPNEYPNGVTYPLSYNDINNA---LN 421
Db 279 RSLNDLTPKALLRKRRLKRLSPNTIKQOYQOQAVNSRNNSSN--SGSTNNMSSMTN 336
Oy 422 ELNSFGDLINPDYTKEPSKNITYTNEKKFTIETKEIKIEKK-----KIESDK-KSYE 475
Db 337 TNNGQSSMAAPNDQY-----NGIYIONLOEFHITNAMEGILNLQKGLKHROYASTKMNDFS 392
Oy 476 DRSKSLNDITKEYEKLEBIYOSKFNNNIDLTNFEKM-----M 513
Db 333 SRSHTIFITTL-YKKHOBELFRISKMLVDLAGSENINSGALNORAKAGSINOSLLTL 451
Oy 514 GKRSYKYEKLTHTNTFASYSKSNHLEKTKALKYMEDYSLRNIYVEKELKYNNLSK 573
Db 452 GRVINALVDKSGH-----IPFRES-----KLTRLQO-----DLSGN 483
Oy 574 IENEIETLVENIKKEBOL-----FEKKTIDENKPKDEKILEVSIYVQVKVLLMKI 628
Db 484 TKTALIAITSPAKVTSEETCSTLEVASAKAKNNKPOLGSFTMKI-----LVKNIT 535
Oy 629 DELKKTQILKVELKHNHIV-PNSYKOENKOEYLLIVLKEIDKLVMPKVESLINE 687
Db 536 MELAKIKSDLLSTKSKESGIYSQDHYKNLNSDLESY-----KNEVOECK---REISLSK 588
Oy 688 EKKNKTEGSDNSEPSTEGEITGOATTKPGQAGSALGDSVOAQOQOAPVPVP 747
Db 589 NALLVKDKLKSKEITQSONCOIESLKTIT-----DHLRQQLDKQKTEIEI--- 634
Oy 748 VPEAKAQPPTPPAPVNNKNTENVSKL-----DYLEKLEFLNTSYICH--KYIIVSHSTM 799
Db 635 -----SDFNKNTLOKLTETVOMALHDYKRRELD-LNOKFEMHITKEIKLKSTL 681
Oy 800 NEKILKOYKTKKEESKLSGCDPLDLFNIONNIPVMSMFDLSLNSLSQLEFMEIYEREM 859
Db 682 FLQI-----NTMOQESILOET-----NIQPNL-----DMIKNEVLTLMRTMQEKAE 722
Oy 860 VCNLYKLDKNDKIKNLEAKKVTSTYKTLSSSMQPLSLTPQDKEPVSA--DQTSHT 917
Db 723 L--MYK-----DCVKILNESPFENVV-----IEKIDILRVDPQKFKYKNAEMLSDIS 769
Oy 918 NLNNSLKLEFENILSGKN---KNIYOELIGOKSSENFYEKILKSDTYNESFTPVSK 974
Db 770 EENNMMKOYL-----KNHFFKNHQBELL-NRHVSTYENIEKRT---NEVEVFKVYL 818
Oy 975 ADDINSLNDESKRRKLEEDIKLKKTQILSFDLYNKYKLELRLFDKKKYVGK-----YK 1029
Db 819 NDHL-----DENKKLIMHMLTTATSAVIDQEMDLFEPRKRVKENSFDLINDCDSMNEFYN 874
Oy 1030 MOIKKLTLEKQLESKLSLNNPKHVLONFVFFNKKKEADIAETENT---LENTKILK 1086
Db 875 SMAATLSQIKSTVDTSSNSM-----ESISVM-----KGOVESENAISLKNNTKFN 923
Oy 1087 HYKGLVYKYNGESSPLK---TLSEESIQTED-----NYASLENFVLSKLEGL 1132
Db 924 QFEDLINKNHMLKDNIKNSITSTSHITNVDDIYNTIENIMKNYGNKKNATKDEMIENIL 983

```



```
Db 1186 SEE--QKILNDIK-----LEISKSVSQYFMKLE--QKSVIGIKSVDFRNNELF 1236
Oy 1033 -----KKLV-----LKEQESKLNSIN-----NPKHVLONF 1059
Db 1237 LSLPIONLTLPEMYRYFEMLYDIHIGIENKANREFIYSSINLDFLINDERVL-NL 1295
Oy 1060 SVFNNKKKEALAEEMLEN-----KKILKHKKGLVKVYNGSSS 1100
Db 1296 EGLIKTKYVLSLSEHRLNISTSPADISIPLOTICPSITTIKK-----TEYYGHLT 1350
Oy 1101 PLKLTSE-----ESTIEDNVASLENFYKSLKGLKDKNL----- 1136
Db 1351 NAMTVASVKKYDPSNIGAINSIDK--SVSDVPALHTIVEBAKYNLMSWDFYNTHASIM 1408
Oy 1137 -----NLE-----KKK-----LSYSSG-----LHHLAEKVI 1161
Db 1409 DTIAROKHSTNIEFHPQSLLEDRDSKGLSLILYLDGTGGYGGYQKLRHNIDTASTLY 1468
Oy 1162 KKNKYTGSPSENNTDVNNALESYKKFLPEGTDVATVVSSEGSPTLESGO-----PKKPA 1216
Db 1469 QTK-YNDNLKLSNDDF-----FLKTKQRIITNSNELGNRKLKNAOLEVELKDP 1518
Oy 1217 ST-----HVGASNT-----TTS 1230
Db 1519 LTEGILYORRISLLITTEYHSLAQIISFWRYTDPNFGHCDPFSLSLAQALFKINITSN 1578
Oy 1231 QNVDEVDVYIPIFGESEED--YDGLGOVYTBV-----TPSYID-----NLSK 1276
Db 1579 RNFSSLYSGGLVKIYFESLSNMKRYIKLPVQTSLSLRLDIYLTPEKLSGSGSLNIMGH 1638
Oy 1277 IENEYVLYLPLAGVYR---SLKKOLENVMTFNVN-----VK 1312
Db 1639 L---VPYSFIDIGVINGNINSESTVKKIRSKIKINDLLOHYINHYISEQTOXIK 1695
Oy 1313 DILNSRKNRENFKNVLESDLIPYKDLTSSVYV---KDPYKFLNKKRDKFLSSVYI 1368
Db 1696 DIYVFLGIQDNTIKYKLESIDIKPISEIQOPLHSILSRQEHVKNLLSGLDSEFSNKLKQ 1755
Oy 1369 KDSIDTQINPANDVLGYKKIKLSEKYSKSDLSIKKYINDKOENEKYLPFLNNIEPLYTV 1428
Db 1756 GLSLTKTVLSVYNNF-----KESKINSIDYEVYTVDLQV-----LYRV- 1793
Oy 1429 NDKIDLEFVILHLEAK---VLNRYTEKSNEYV---KIKELANLYKTQDCLADFKKNNFVG 1481
Db 1794 --DIDTVIGIGTFKEGINSLSLEAHNNIDAINSGVIGVYQARMI-----KANDNSA 1844
Oy 1482 IADISTYNNHNNLTKFLSTGWFENLAKTYLSNLLDGNLQGMUNISOHCYK 1535
Db 1845 IDHAGAVSDIKNIYDKFLGILTLTNRVNPNPGVSGASLEGTSSGLEVCASR 1898

RESULT 23
US-08-290-919-3
Sequence 3, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DAREY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714637 CUSH
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 53 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label=X
OTHER INFORMATION: /note="X = M and N, or N"
```

```
US-08-290-919-3
Query Match 3.6%; Score 301; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 7e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1570 PNPCCNENNGGCDADAKCTEDSGSNGKIKICECTKPPSPYLPFOISFSSSN 1621
Db 2 PNPCCNENNGGCDADAKCTEDSGSNGKIKICECTKPPSPYLPFOISFSSSN 53
```

```
RESULT 24
US-08-480-604A-6
Sequence 6, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHAY, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

Query Match 3.5%; Score 298.5; DB 1; Length 2710;
Best Local Similarity 20.4%; Pred. No. 1.2e-07;
Matches 339; Conservative 262; Mismatches 573; Indels 487; Gaps 90;

QY 84 SGGSVASGSGVASGSGVAGSGGNSRRTPNSDSDAKSVADLKHRRVANYLTITK--E 141
DB 538 TGGSLSEDDGCVDFNKNTALDKNYLLNNKIPSNVVEBAGSKNYV-----HYIIQLQGD 590
QY 142 LKYPQLFDL-----TNHMLTLCQNHGFK--YLIDGYEELNELLYKLNFFYDILLRKIND 194
DB 591 ISTEATCNLFSPKPKNSITIIQRNMNESAKSYFLSDGESTLEL----- 633
QY 195 VCANDCQIPENLIRANELDVLYLKVFGY-----RKPLDNKDNVGMEDYIK 243
DB 634 ---NKY-RIPERLKNK-----EKVYVFIGHGKDEFNTSEFARLSVDSLSEISFLDTIK 685
QY 244 K--NKRTIE-----NIN-----LIESKKTIT-DKNKNATKEEKKKL 278
DB 686 LDISPKNVEVNLGCMNFSYDFNVEETYPGKLLSLMDKITSTPLPVNKNKSI-----TI 739
QY 279 YQAOVYLSYNNKOLEE--AHN--LISVLEKRIDTLKKNENIKELDKI--NEIKNPAPNS 333
DB 740 GANQYEVRIINSEGRKELHASKWIKKEBAJMSDLSKEYIT--FDSIDKKL-----AKS 793
QY 334 GNP-----NPLLDKNKKIEHEKEIKIARTIKFNIDSLFTPLLELYLREKN 383
DB 794 KNIPGLASISEDIKTLILDASY-----SPDTKFTILNNKLINSSIGDYI--YY----- 840
QY 384 KNIDISAKYETKESTEPNEYPNGVTPPLSYNDINNALNELNSGDLINPDYIKESKNI 443
DB 841 -----EKLEP-----VKNIH--NSIDDLIDEFNLEENYDEL 871
QY 444 YNDNEKRFINEIKIEKKIESDKKSYEDR-----SKSLNDI 484
DB 872 Y---ELKK-LNNIDEXYLISFEDISKNNSTYSVRFINKSGESVYETEKEIFSKISEHI 927
QY 485 TFEYERKLNIYDSKFN--NIDLTNFEKMGKRSYKVEKLTHTNTPASYSKHNIE 541
DB 928 TREISITIKSIITDVGNNLLDNTQDHTSOVNTLNMAFLQSL-----IDYSNNMDVLA 981
QY 542 KLTALK-----YMEDYS--LKNIVYERKLYKKNLISKIENE-----IETLV 582

DB 982 DLSTSVKQVOLAQLFSTGLNTIYDSIQL---VVLISNAVVDITNLPTTEGIPIVSTIL 1038
QY 583 ENIK-----KDEQLFEKKITKDEKPDKILEVSDIVVQYQKVL----- 623
DB 1039 DGINLGAATKELLEDHDPILKREL---EAKVGLAINMSTISIAATYASIVGIGAEVTIPL 1095
QY 624 ----LMMKIDELKKTQLLIKNLVNLHNNHVPNSYKQENKQEPYLLVLKKEIDKLKVEF 678
DB 1096 LPTAGISAGIPSLVNNELIHD--KATSVVYFNNHLSKSKYGP-----LKTEDD--KILV 1147
QY 679 PKVESLINEKKN-----ITBEGSDN-----SEPSTBEITGQATTKP 717
DB 1148 PIDDLVISEIDFNNNSIKLGTCNILAMEGSGFTVGNIDHFPSSPSISHI-----P 1200
QY 718 GQOAGSALEGDSVOAOAOBQKQAPVPVYVPPAKAQQ---PPPPVNVNKTEN--VSKL 772
DB 1201 SLSTYSAIGIETENLDPKSK-----IMMLPAPASRVFMEWEGAAPGLRSLENDGTRLL 1253
QY 773 DYLEKLY--EFLNTSYCHYYILVSHSTNNEKILKQYKITEEESKLSSCDPLDLLFNIO 830
DB 1254 DSIRODLPCKFYWRFYAFEFYAITTLKPYEDTNIRIKLIDKDYRNFIM--PTITTEIR 1310
QY 831 NNIPWMSFDSLNSLSQLFMEIYEKEMVCNLYKLKDNKIKNLLLEAKVSTSVKTL 880
DB 1311 NKL--SYS-FDGAGGTYG--LLSSSY--PISTNINLSKDDMLPINIDNEVEAREISENGTJK 1364
QY 891 SSSMQLSLTPQOKPEVASADOTSHSTNLNLSLKLENITIS--LGKNNIYQELIGOKS 947
DB 1365 KG-----KLRKDVLSKIDINKKLI-----IGNOT 1389
QY 948 SENFYEKILKSDPTFYNESTNPFVKSKADDINSLANDESKRKLLEDINKLKTLOLSFDI 1007
DB 1390 ID--FGSDIDNKRXYI-----FLTCELDKISLIIE-----INLAVKS----- 1425
QY 1008 YNKYKIKLELPLPKKTKTVGKYKMOIKKLLKQLESKILNLPKPHVLONEVFPNKKK 1067
DB 1426 ---YSLLSG--DKNYLISMWNTIEKINTL--GLDS-----NIAIVNYDESNNK 1470
QY 1068 EAEIATENTLENTKILKLY-----KGLVYKYGESSPLKTLSESIQFEDNYSLENFK 1123
DB 1471 FGAIKST-----SOKSIT--HYKDSKNILEFYV--DSTLEFNKSDPLAEDINVPKMD-- 1520
QY 1124 VLSKLEGLKLDNLEK--KLSTYS-----SGLH--HLIAELKVIKKNKYGNSP 1171
DB 1521 -INTIGKYYVDNNTQKSIDFISLVSKNQVKNGLYLNESYSSYLDEFKKNSDGHNNIS 1579
QY 1172 SENNTDVNNALESYKKEFLPEGTQVA-----TYVSESGSDLE--QSQPKRPASHVGAESN 1225
DB 1580 NEMNLEFLDN--ISFWKLGFEFNINIEVIDKYFTLVGKTNLGVFEICONNKNIDIFYEGEMK 1638
QY 1226 TITTSQNVDEVDVLIIVPFGESEEDYDGLQGVNVEAVTPVIVINILSKIEENEVLY 1285
DB 1639 SSKSKTIFSGNGNVYVEEPLNPD-----IGEDISTG-----LDFSTE-- 1676
QY 1286 LKPLAGVYSLKK-----QLENNVMTFNVNVKD-----ILN--SPKNRENFKNVLS 1331
DB 1677 --PLYGIDRIKINKVLIAPDLYTSLINININNYSENYEPEIIVNPPYTFHKVWIN--LDS 1732
QY 1332 DLIPYK-----DLTSSNVVKKDPYKFLNKEKKDKLSS--SYNLIKDSID--TDINPANDVL 1383
DB 1733 SSFEYKMWSTEGSDPLIVRVLSESNKKILQIKIRKGLISLTQSNKMSIDPKDKIKLSS--L 1790
QY 1384 GY-----YKILSEYKSDLDISI-----KKYINDKOGENEKYLPLFN-----NI 1421
DB 1791 GYIMSNFKSGNSNENELDRHGLGFIIDNKTYIYDEDSKVLKGLININNSLFIYDPTEFNL 1850
QY 1422 EPLKYTVNDKIDLEVLHLEKVLNYYTERKSNEVAKIKELNLYKTIDOKLADFEKNNPVG 1481
DB 1851 VTGQWOTINGKKYFYDINTGALTSYKI-----INCKHFEFNDGVQWOLGVFPGDGFEX 1904
QY 1482 IADLSTDYNNHNL-----LTKFLSTG--KVFENLAKTV 1512


```
Db 1201 SLSTYSAIGIETENIDFSK-----IMLFPNAPSRYFMETGAVPGLRLSLENDGTRL 1253
Qy 773 DYLEKLY--EFLNTSYICHKYLIVSHSTMNEKILQYKITEEESKLSODDLFLFNQ 830
Db 1254 DSIRDLPGKRYWRYAFDAITTLKRYEEDTNIKIKLDKOTRNFIM--PTITTNETR 1310
Qy 831 NNIPWYMSFDSLNSLSQLEMEIYKEKEMVCLYKLDKNDKIKNLLEAKKYSTSVKLS 890
Db 1311 NKL--SYS-FGAGGTSY-LLLSY--PISTMINSKDOLMFINDNEVREISENGITK 1364
Qy 891 SSSMOPSLTLTQDKPEVSANDTSHSTLNNSLKLEPNILS--LGNKNKYOEILIGOKS 947
Db 1365 KG-----KLKIDVLSKIDINKKLI-----IGNOT 1389
Qy 948 SENFEKILKSDPFYNESFTNFVSKADDIINSLNDESKRKLEEDIKKLTLOLSPDL 1007
Db 1390 ID--FSGIDKDKRY-----FLCELDKSLITE-----INLVAS----- 1425
Qy 1008 YNKYKLELRLERDKKKRYGKQJIKLTLLEQLESKLNLSLNPKHVLONFVFNKK 1067
Db 1426 --YSLLSG--DKNYLSNLSNTEKINTL--GLDS-----KNLAYNTDESNNKY 1470
Qy 1068 EAEIETENTLENKILKHY--KGLVKYNGESSPLKLTSESIOTEDNYASLENPK 1123
Db 1471 FGALSKT-----SOKSII-HYKKDSKNILEFYN--DSTLEFNSKDFIADIVFVKDD-- 1520
Qy 1124 VLKLEKGLKDNLEK--KLSYLS-----SGLH--HIALEKYEIKKNNTGNSP 1171
Db 1521 -INTTGKYYDNDNTDKSIDISILVSKNOYKNGILNLESYSSYLDOPKNSDGHNTS 1579
Qy 1172 SENNTDNNALLESYKFLPECTDVA--TVVSESGSDTLE--QSQPKPASTHVGASN 1225
Db 1580 NEMNFLDN--ISFMKLFGEFENINFIIDKYPFLVGTNGLGYEFIDGNKNKIDYGEKMT 1638
Qy 1226 TTTTSQNVDEVDYIYPIGSESEYDGLQVVTGCAVPSYDNLISKTENEYEVLY 1285
Db 1639 SSSKSTIPSGNGRNVVEPIINPD-----TGEDIYSTS-----LDPSYE-- 1676
Qy 1286 LKPLAGVYRSLK--OLENNVMTFNVNVKD-----ILN--SREKNRKNFNVLES 1331
Db 1677 --PLXGIDRIYINKYLIADLTSLNINTNYNSNEYPEIILYLANNTTHKKANNIN--LDS 1732
Qy 1332 DLIPYK-----DLTSSNYVYKDPFLNKEKROKFLS--SYNYIKDSID--TDINPANDV 1383
Db 1733 SSFEKMGTEGSDFLVRYLESNNKLIQKIRKIGILSNTQSPNMSIDFKIKLS--L 1790
Qy 1384 GY-----XYLSEKYSKSDLSI-----KTYINDKOGENEKYLPLN-----NI 1421
Db 1791 GYMSNFEKSPENSEMELDRDHGFKIIDNKTYYYDEDSKLVGLININNSLFFYDFIEFNL 1850
Qy 1422 ETLKTVNDKIDLFYIHLAEKLVANTYKESNVEYKIKELANTYKIDODLADPFKNNNVG 1481
Db 1851 VTGMOTLNGKYYFDINIGALTSYKT-----INGKHFYFNDGVMOLGVFKGPDGEY 1904
Qy 1482 IADLSTDYNNNL-----LTKFLSTG--MVENLAKTV 1512
Db 1905 FAPANT--QNNNIGQALVYOSKFLTLNGKRYFEDNNSKAV 1943

RESULT 27
US-08-957-310-6
; Sequence 6, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
```

```
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-957-310-6

Query Match 3.5%; Score 298.5; DB 4; Length 2710;
Best Local Similarity 20.4%; Pred. No. 1.2e-07;
Matches 339; Conservative 262; Mismatches 573; Indels 487; Gaps 90;

Qy 84 SGGSVASGGSVAGSGVSGSGSRRTPSPSSSDSAKSYADLKHRYNLYLTK--E 141
Db 538 TGGSLSENGVDVFNKNTALDKRYLLNNKIPSNVVEAASKYV-----HIIDLGDD 590
Qy 142 LKYPFDL-----TNHMLTLCDNTHGFR--YLIDGYEINELLYKLNFPYDILRAKLN 194
Db 591 ISEATCNLFKSNPKNSIIIOQNNMESAKSYFLSDGSEITEL----- 633
Qy 195 VCANDYCOIPENLKRANELDYLKALVGY-----RKPLDNKDNVGMKEDYIK 243
Db 634 --NRY-RIPERLKNK--EKVKYTFIGHGKDEPNTSEFARLSYDLSLSNELSFLDIK 685
Qy 244 K--NKKTIE-----NINE-----LIESKKTII--DKKNATKREKKKL 278
Db 686 LDISPKNVEVNLGCMNFSYDPNVEETYPGKLLLSIMDKITSTLPDVKNNSI-----TI 739
Qy 279 YQAGYDLISYKNOLEE--AHN--LISVLEKRIIDLTKKNEIKELDKI--NEIKNPPANS 333
Db 740 GANQVEYRINSEGRKRELLAHSGKWINKREALMSDLSKYEYI--FEDSDLNKIK--AKS 793
Qy 334 GNTP-----NTLIDKNNKLEEHKEIKETAKTIKFNIDSLFTDPLELEYLYREKN 383
Db 794 KNIGPLASISDITKILLDASV-----SPDYKFIUNNKLNISSIGDYI--YY----- 840
Qy 384 KNIDISAVKRESEPEPEYNGVYPLSYDNNALNELNSFGDLINPPTYKPSKNI 443
Db 841 -----EKLEP-----VKNIH--NSIDLDIDFNLLENVSDGL 871
Qy 444 YTDNERKKFINKIKIEKKIKIESDKSYEDR-----SKSLNDI 484
```

```

Db 872 Y---ELKK-LNNIDEXLLSFEDISKNNSTYSVRFINXNGSVYETEKEIFESKSEHI 927
QY 485 TREYEXELNIDYXSKFN---NIDLNFEMKMKRYSYXEKJLJHNTFASYSKNHLE 541
Db 928 TKRISTIKNSITIDVNGNLDNIQDHTSOVNTLNAFFIQSL-----IDYSNNKQVLN 981
QY 542 KLTAKLK---YMEDYS-LBNIVYERELKYYKLLSKISENE-----LETLV 582
Db 982 DLSTSVKVOYADLFSTGLFTIYDSIOL---VNLISNAVDJTNVLPTEIGAPIVSTIL 1038
QY 583 ENIK-----KDEQLEFEKKTIDENKPKDEKILEVSDIVVQOKVL-----623
Db 1039 DGINLGAIKELLDHEDHPLKREL---EAKVGLAINMSLSIAATVASIIGAETIPL 1095
QY 624 -----LNNKIDELKKTQIULKJNVELKHNHVPVSYKOENQOEPUYLVLEKKEIDKLKVF 678
Db 1096 LPLAGISAGIPLSVNNEILLHD-KATSVVYFPHLSKRYGP-----LKTEDD-KILV 1147
QY 679 PKVESLINEEKN-----IKTEGOSDN-----SEPSTEGLITGOATTKP 717
Db 1148 PIDDLVISEIDFNNSIKLGTGNILAMEGSGHTVTGNIDHFFSPISISHI-----P 1200
QY 718 GQAGASLBEDSVQAQOEOKAQOPPVYVPEKAKY---PPPAVYNNKTEH--VSKL 772
Db 1201 SLSTYSAGIETENDLSK---IMLPNAPSKEVFWETGAVPGLSLEDGTTRL 1253
QY 773 DYLEKLY--EFLNTSYICHYIIVSHSTNNEKILKQYKTEBESKLSGCDPLDLEFNIQ 830
Db 1254 DSJRDLYPGKFEYWRFAFPFYAITTLKPYVEDNINIKIKLKDORNFIM---PITTTNEIR 1310
QY 831 NNIPVATSMDSLNSISJOLFMEIYEKEMVNCYLKLDNDKIKNLEAKKVSIVKTL 890
Db 1311 NKL--SYS--FDGAGCTYS--LLLSY--PISTNINLSKDDLMFINIDNEVREISEIENGTIK 1364
QY 891 SSSMOPSLTRPOKPEVSANDTSHSTNNLSKLFEINIIS---LGKNKNIYELQIGKS 947
Db 1365 KG-----KLKDVLSKIDINKKLI-----JGNOT 1389
QY 948 SENFEYKILKSDTYNESEFTNFKVSKADINSINDESKRKLLEEDINKKTLQLSFDL 1007
Db 1390 ID--FSGDIDNDRYI-----FLTCELDIDISLIE-----INLVAKS-----1425
QY 1008 YNNYKLELEFDDKKTGVGYYKMQIKKLTLLEKDESKLSLNNPKVAVLONFSVFFKKK 1067
Db 1426 ---YSLLSG--DKNYLISMJMTIEKINPL--GLDS-----KNIAVNYDESNNKY 1470
QY 1068 EAEIETENLEMTKILKKHY---KGLVAYNGESSPLKTLSEESIOTEDNVASLENFK 1123
Db 1471 FGAISKI-----SOKSII-HYKDKSKNILEFYN--DSTLEFNSKDFIAEDINVFMRKD-- 1520
QY 1124 VLSKLEGLKLDNLEK---KLSYLS-----SGLH--HLIAELKEVIKNNKYTGN 1171
Db 1521 -IMTYTKYVYVNDNTDKSIDFSISLVSKNOVKVGLYLNESVSSYLDFAKNSDGHHTS 1579
QY 1172 SENNTDVNNALLESKKRLPBGTYA-----TVVSESGDPLE-QSQKKRKAASHVGAESN 1225
Db 1580 NFMNLFIDN--ISFWKLEGFENINEVIDKYFTLVGKTLGVEFTCDNNKNIDILYFGEWKT 1638
QY 1226 TITTSQWVDEVDVILVPIFGESEEDYDGLGVKVTGEAVTPSYVIDNLSIKENEYVLY 1285
Db 1639 SSSKSTIFSNGRNVVEPIYNP-----TGEDISTS-----LDSTYE---1676
QY 1286 LKPLAGYRSLK---OLENNVMPFNWVKD-----ILN-SFNNKRENEKNVLES 1331
Db 1677 --PLYGIDRIINKVLPDLTYSLININTNYSNEYPELIIVLPNPFHKKVNIN--LDS 1732
QY 1332 DLIPYK-----DLTSSNVYVADPYKFLNKKRDKFLS--SYNTIKDSID--TDINFANDVL 1383
Db 1733 SSFEYKSTEGSDPILVRYLEESNNKILQKIRIKGLISNTQSFNKSIDKDKIKLS--L 1790
QY 1384 GY-----KILSEKYSKSLDSI-----KTYINDKQGENEKYLPFLN-----NI 1421
Db 1791 GYIMSNKSPNSNEMELDRHDLGFKRIIDNKTYVYDEDSKLVGLININNSLPFDPPIEFNL 1850

```

```

QY 1422 ETLKYTVNDKIDFVHLEKVLNTYREKSNVEVKEIKELNYLKTIODKLADFKNNFVG 1481
Db 1851 VTMQOTINCKKYYFDINTGALALSYKI-----INKKHFEFNDGVWQGLVFGPDPGEF 1904
QY 1482 IADLSTDYNNHNL-----LYKFLSTG---WVEENLAKTV 1512
Db 1905 FAPANT--QNNNIEGOAIYQSKFELTNGKRYFEDNNSKAV 1943

RESULT 28
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf5230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170B
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ouline, Jonathan, A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 015280-113100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ. ID NO.: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-323-170B-2

Query Match 3 5%; Score 298; DB 1; Length 3135;
Best Local Similarity 19.6%; Pred. No. 1.5e-07;
Matches 347; Conservative 270; Mismatches 573; Indels 582; Gaps 92;

QY 128 LKHVRVNYLLTIRE---LKYVPLDFTLNMJLTCDNHGFKYLID---GYEINELLYK 180
Db 1502 LKNNLANF--TYKHQFYVMILPALMDNDISFKCYVDLKKKKNVNSPLGPYVLRALYKK 1559
QY 181 LNFYFDLLRAKLDVNCAND---YCOIPFNLIKIRANELDVLAKLV--FGYKRPIDNIRKDNV 235
Db 1560 LNIKFD-----NYVTGTDQNKYLTMYMDHL--SHKKNYLLKELFHDGKKKPAD-----1606
QY 236 GKMEDYIKKKKTIENINELIESKKTIDKNNKA---TKREEKKKLYQAOYDL--SLYNN 290
Db 1607 -----TDANESII--ESLSINESNESGPFPTGDVDAEHLILEGYDTWESLYDE 1653
QY 291 QLEEA--HNLISVELEKRDITLAKNNEIKELDKINEIKNPPRANSNTPTNL-----LDKNN 345
Db 1654 QLEEVYINDIESLEK-----DIEQVYVLA--LKAPKLMSAQIHNHRHVCDEFSKNN 1704

```

Qy	346	KIEHEKEIEIKIAKTIKFNIDS-LFDPDLLEYLLAEKKKNIDISAKV-----	392
Db	1705	LIVPESLKKEBEGNPNVNIHYALLKPLDPLVLCOPTSDKDYAEAKVNISENDNEYELO	17644
Qy	393	-----EKEESTEPNE-----YPNGV--TYPLSYNDINNALNELNSGDFILNFD	434
Db	1765	VISLIEKRPHNFETLEBSKPGPDGVVHNGVDTPGV-----LDNSPFEKYFNIKIRK-D	1819
Qy	435	YTERPSKNITDNEKKFPIE-----IKERKITEK-----KI	467
Db	1820	KFEKVIYNEDDTEEEKDESILPGAIVSMPVKLKKRDPFTSYAFAVPPIVPKDLHKV	1879
Qy	468	ESDEKSEYDESKSYLN-----DIRKEKELNLEYDSEFNNNIDLFNEKCM-GKRSY	518
Db	1880	ECNNTEYKDENOYISGNIIGHIDISNKRING--CDSSTNNSSLTSSVKLVNGEPIKN	1937
Qy	519	YKV-----EKLHNHTASYENSKNLEKTLKAKYMEDYSLNI	558
Db	1938	CEIINNNEVFGIICDNETNLDPEKCFHE---IYSKDKVTKYKPREVIJPNIDPSLHN	1992
Qy	559	VEKELKYK--NLISKI-----ENELTELVENIKDEEO-----	591
Db	1993	SNKKRVAYAVPDLIYNKLLFSCSCKTSHNTIIGTKVITLMDKEEEDFKTAOGIKHN	2052
Qy	592	-----LEEK-KITKDEKNPDEKILEVSDIYKVOVKVLLMKINDELKTKOUL-----	638
Db	2053	VHLCNFPDNLFPDNNK-----IYLCITDLELSEVILOLPIGOT	2093
Qy	639	KNVELKINIHYPNYSKOENKOEPIYILVLEKKEIDKLVMPRVESLINEKKNIKTGEO5	698
Db	2094	KNVE-----EGVONEBY-----KFKSLRPSLVFDDNNNDIKVIKE	2129
Qy	699	DNSEPTEGETTGATTKPQOAGSALDEGSOVQAOQEOKO-----PRVPVYVPAK	752
Db	2130	KN-----EVSISLAK-----GYGKRITPFRKNKKKEGSIPLPI-----	2167
Qy	753	AQVEPPAPVNNKTENVSKIDYLEKYLEPLNTSYICHKYILVSHSTMEKKILKOYKITE	812
Db	2168	-KODTDLKFIINETIDNSNIKORGLIYIVRWKVSENSEFKLDDFTTGSTISMEILNSOYKE	2266
Qy	813	EES--KUSCDPLDLFPIONNNIPWVWSMFD--LUNSLISOLFMELYEKEMWCNYKLKD	868
Db	2227	KKCYVAKIKKGD---IFGLK--CPKGFALFPQACSNNVLEYKSKDYDESEHINITYHKD	2280
Qy	869	---NDKIKNLEBAKVVSTSVKTLSSSSMOPLSTPQDKREVSANDDTSHSTNIN--NSLK	924
Db	2281	KKYMLKPKOYIE-----LMDENPRELO-NIOQYTGISNTIDVLHFNENMLGLRP	2328
Qy	925	L-----FENISLGRN-----KNYQELIGKSSDENFYEKTLKDSOTFY	963
Db	2329	LNFKNHYSTAVAKVPDTFNSIINFSCNCPNPEKHVHYGTQVSDNRNF-DNKKKNENYK	2387
Qy	964	NESFTFNFKSKADDINSLNDESRRKLY-----EEDINKEKTKLOLSFDLYNNKYKLK	1018
Db	2388	NFLPLNIEKAL-----LLDEEROKKIKIOQOEEQEOJOLKODDRLSRHDDYKNNHITYI	2443
Qy	1016	ERLEPDKKTYGKTYKIOKIKITLL--KEOLESKLSLN-----NPKHV	1055
Db	2444	--LIDSNEHICDEYKESNLSTLPNDTKIOKSICKINKAKALDVYTIKCPHTKNFTPEDY	2501
Qy	1056	LONSVPFNKKKEAEIATENTLENTKILKHNKGIVKYUNNESSPLTKLTPSESIQOTEDN	1115
Db	2502	FPNSSLITNDKK-----IYTFDKKNRYVYIDPTK-----KTFSLKDIYIOSF	2544
Qy	1116	Y-ASLENFVYLSKLEGLKLDNLNLEKKKLSYSSGLNHLIAELKEVINKNKNTGNSPSEN	1174
Db	2545	YGVSLIDLHANOIKIKIHEEMWD-----VHLFEPRHV--LHHVVYLNHHI-----	2584
Qy	1175	NTDVNNALLES--YKKFLPESTDAVATVYSSGSDTLEQSPKPKPATSHNGAESNTITTSQN	1232
Db	2565	-VNLSASLBEGVLEFKSKVTDDEAT-----KKNYTLPT--	2616

```

01 1233 VDDDDVIVIPICESEDEDDLGQVYTGAVT--PSVIDNIIISKIE-----NE 1280
02 2617 --DGVSIIILIPYKE-----DIFHLFCGSGTCKPKKNTSALIHIISSNNTIIG 2669
03 1281 YEVLKPLAGYRRLKQLENNVMTFPVNVKDL-----NSRFKRE 1323
04 2670 CDFLELQENQNDALISNNNNNSYSIFTHKNTENNLCIDISILPKTVIGIKCPKNLPQT 2729
05 1324 NEKNV--LESDLIPKDLTSSNY-----VVKDPYKFLNKKRDKFLSSNYVT-- 1368
06 2730 CFDEYVYVQEDVPKSTITADKYNFPFSKODKIGNILKNAISINNPDEKD--NTYTYLILP 2786
07 1369 ----KDSIDTDINPANDVLGYKYLISEKYSKSDLSIKAYINDKOGENE--KY-----LP 1416
08 2787 EKFEELIDTRKVLACTGDKNYIIHMKIEKSTMDKIK--IDEKKTIGDICKYDVTTRVA 2844
09 1417 FLNNETLYKPV-----NDKI---DLFY--LHLEAKVLNY 1446
10 2845 TCEIITDIDSSVLAKEHHTVHVHSTLSRDKLIILKPTNEKTHFENFVNPENLKKVLY 2903
11 1447 TYEKS--NVE-----VKIKELNYLKTIDOKLADFKKNNEFVGIAIDSTDYN 1490
12 2904 NWNKNIENIHILPGAITTDIYDTRKIQ--YLILIPPYV---HNDIHF-----SLEFN 2952
13 1491 HNHLITKFLSGWMEFENLAKTVYLSMLDGNLQGMNLISQ-----HQCYKQCPONSGCFR 1545
14 2953 NLSLTK--QONQIITYGNVAKI-----FIIHNGYKEIHGC-----DFTGKYS 2993
15 1546 HLDERECKCLNYYQEGDKCVENPNPTCNENNG 1579
16 2994 H-----LFTYSK---KPLPNDIDICNVITG 3015
17
18 RESULT 29
19 US-08-954-441-2
20 : Sequence 2, Application US/08954441
21 : Patent No. 6316000
22 : GENERAL INFORMATION:
23 : APPLICANT: Williamson, Kim C.
24 : APPLICANT: Kaslow, David C.
25 : TITLE OF INVENTION: Cloning and Expression of Plasmidium
26 : TITLE OF INVENTION: falciptatum Transmission-Blocking Target Antigen, Pfs230
27 : NUMBER OF SEQUENCES: 4
28 : CORRESPONDENCE ADDRESS:
29 : ADDRESSEE: Townsend and Townsend and Crew LLP
30 : STREET: Two Embarcadero Center, Eighth Floor
31 : CITY: San Francisco
32 : STATE: California
33 : COUNTRY: USA
34 : ZIP: 94111-3834
35 : COMPUTER READABLE FORM:
36 : MEDIUM TYPE: Floppy disk
37 : COMPUTER: IBM PC compatible
38 : OPERATING SYSTEM: PC-DOS/MS-DOS
39 : SOFTWARE: Patentln Release #1.0, Version #1.30
40 : CURRENT APPLICATION DATA:
41 : APPLICATION NUMBER: US/08/954,441
42 : FILING DATE: 20-OCT-1997
43 : CLASSIFICATION: 424
44 : PRIOR APPLICATION DATA:
45 : APPLICATION NUMBER: US 08/323,170
46 : FILING DATE: 13-OCT-1994
47 : PRIOR APPLICATION DATA:
48 : APPLICATION NUMBER: US 08/010,409
49 : FILING DATE: 29-JAN-1993
50 : ATTORNEY/AGENT INFORMATION:
51 : NAME: Einhorn, Gregory P.
52 : REGISTRATION NUMBER: 38,440
53 : REFERENCE/DOCKET NUMBER: 015280-113110US
54 : TELECOMMUNICATION INFORMATION:
55 : TELEPHONE: (415) 576-0200
56 : TELEFAX: (415) 576-0300
57 : INFORMATION FOR SEQ ID NO: 2:

```


SEQUENCE CHARACTERISTICS:
 LENGTH: 3135 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-954-441-2

Query Match 3.5%; Score 298; DB 4; Length 3135;
 Best Local Similarity 19.6%; Pred. No. 1.5e-07;
 Matches 347; Conservative 270; Mismatches 575; Indels 582; Gaps 92;

QY 128 LKRVNRYLLTITKE-----LKYPOLFDLTNMLTLCNINHFKKLID--GYEINELTK 180
 DB 1502 LKNNLNF--TYHQFNMEIPALMDNDISFCICVDLKKKTKYKSPGLKRLALYKK 1559
 QY 181 LNFYDLRLAKLNDVAND--YCOIPFLKIRANEDLVKKLV--FGYKPLDNIKDNV 235
 DB 1560 LNKFPD-----NYVTGTOONKLYMTYMDLHL-SHKRNLYKELFLHDLGKKKPAD----- 1606
 QY 236 GKMEYIKKNNKTITENINELIESKKTIDKNKA---TYEKKKKLYQAOYDL--SIYKN 290
 DB 1607 -----TDANPESTI--ESLSINESNESGPPPTGVDAEHLILEGYDTWESLYDE 1653
 QY 291 QLEEA--HNLISYLEKRIIDLKKNENIKELIDKINELIKNPSPANGMPTNL-----LDKNK 345
 DB 1654 QLEEVLYNDLESLELK-----DIEQYVLOVN-LKAPLMSAOIHNNRHVCDPSKNN 1704
 QY 346 KIEEHKEIKELAKTIKFNIDS-LFTDPLELEYLYREKKNKIDISAKV----- 392
 DB 1705 LVIPELSKKKEELGKNPVNHCYALLKPLDTLYVCPTSKDNEEAKVNSENDNEYELQ 1764
 QY 393 -----ETKESTEPNE---YPNGV--TYPLSTVDINNALNELNSFDLNPED 434
 DB 1765 VISLEKREHNEFTLESKRKRGNGVYVHNGVDTGPV---LDNSTEKKYFKNKKIKIP-D 1819
 QY 435 YKPEPKNIYTDNERKKFTINE-----IKERIKIERK-----KI 467
 DB 1820 KPEFEKINIEDOTEEKOLESLIPGALVSPMKYLKKKDPPTSTAATVVPPIYVKDLHFYK 1879
 QY 468 ESDKSKYEDRSKSLN-----DITKEYEKLINELIYDSKFNNDLTNEKKM--GRRYS 518
 DB 1880 ECNNTFEKDNQYISGVNGIHIDISNSNRKING--CDFSTNNSILTSVKLVNGETKN 1937
 QY 519 YKV-----EKLTHNTPASYSKSHNLEKLTAKALYMDYSLRNI 558
 DB 1938 CEININNEVEFGICDNEINLDPKCFHE---IYSKDNKTYKRFREVIPIINDIFSLHN- 1992
 QY 559 VVEKELKYK--NLISKI-----ENETPLVENIKKDEO----- 591
 DB 1993 SNKKKAYIAKVPDIYINKLFCSCSCTSHNTITGTMKVTLNKKDEKEDEDEKTAQGIKHNN 2052
 QY 592 -----LFEK-KITDENKPDDEKILEVSDIVKVOVVLNKNIDELKKTOLIL----- 638
 DB 2053 VHLCFNFDPELFDNNK-----IYLCIKIDAEFLSVIILQILFPG 2093
 QY 639 KAVELKHNHIVNSYKOEKQEPYLYLVKKEIDKLKVPKPKYKESLINEKKAITKEGOS 698
 DB 2094 KNE-----EGVQNEEY-----KKESLKPSPVLEDDNNNDIKVIYIGKE 2129
 QY 699 DNSEPTGEITGOATTKPGQOQGSALLEGDSYQAOQOQAO-----PPVVPVPEAK 752
 DB 2130 KN-----EVSISLALK-----GYGNRIETFPKNGKKGGISFFIPLI----- 2167
 QY 753 AQVPPAPAVNNKTEVNSKIDYLEKLEFNTSYICHKYLIVSHSTMEKILKQYITKE 812
 DB 2168 -KODTDLKFIINFTIONSNIKQGLIYFVRKVVSENSFKLDCPTGTSISMLNQVKE 2226
 QY 813 EES--KLSSCDPLDLFNIGNNIPVMSFDS--LNNLSQLEFMEIYEREMVCNLYKLD 868
 DB 2227 KCTVYIKIKGD---IFGLK--CPKGFALFPQACFSNVLEYYKSDYEDSEHNIYIHKD 2280
 QY 869 ---NOKIKMLLEBAKKVSVTKLSSSQOPLSLTPQDKREVSANDTSHSTLNL-NSLK 924

DB 2281 KRYNLKPKDVI-----LMDENPRELO-NIQOYTGISNTDVLHFKNNENIGLNP 2328
 QY 925 L-----FENILSLGN-----KNIVOEILIGOKSSNPEFKILKDSDFY 963
 DB 2329 LNFKNHYSTAYAKVPDTPFISINFCSCNCPNPEKHVYGTQOVESDNNNF--DNIKKNENYIK 2387
 QY 964 NESFTNFKSKADINSLNDESKRRKL-----BEDINKLTKYLQSLFIDLYNKKYKL 1015
 DB 2388 NFLPLPIEYKAL-----LLDDEERQKKIKOOQEEQOEOQLIKDDRLSRHDDYKNKHTYI 2443
 QY 1016 ERLFDKKTKYGVKKMQIKKLTLL-----KEOLESKLSLN-----NPKHY 1055
 DB 2444 ---LYDSEHICDYEKESLSLTPNDTKIKIOKICKIMKALDVTYIKCPHKNTPPKDY 2501
 QY 1056 LQNFVSEFKKKEAEIETENLTENTLILKHYKGLVYKYNGESSPLKTLSEESITQEDN 1115
 DB 2502 FPNSSLITNDK-----IYATFDKKNVTVYIDPFR--KTFSLKDIYIQGF 2544
 QY 1116 Y-ASLENFVLSLEKGLDNLNLEKKSLSYSSGHHLLAEKYEIKKKNYTGNSPSEN 1174
 DB 2545 YGVSLDLNOLIKKIHEDMD-----VHLEFPNHN--LANNVLNNHI----- 2584
 QY 1175 NPDVNNALSS--YKFLPREGTQVATVVSSESDTLBQSPKPKPASTHGAESNTTITSON 1232
 DB 2585 -VULSSALBGLVFMKSKVYQDETAT-----KKNITLPT----- 2616
 QY 1233 VDDEVDVVIIVPIEGSEEDYDDLGQVTVGEAVT--PSVIDNLSKIE-----NE 1280
 DB 2617 --DGVSSILIPYKVE---DITFHLFCQKSTTKPKNNKNSLALIHHSNRNIIING 2669
 QY 1281 YEVLVYKPLAGVYRSUKOLENNVMTFNVNKOIL-----NSRPNKRE 1323
 DB 2670 CDFLYLENQTDNAISNNNNNSYSIFTHNNTENNLLICDISLPKYIYIGCPKKKLNLPQT 2729
 QY 1324 NFKNV--LESDDLRYKDLTSSNY-----VADPKYFKLKEKREDFLSSYNYI--- 1368
 DB 2730 CDFEVYVVOEDVPSTIYADKNYFMSKDIGNILLNALSINNPDKD---NMYTYLLIP 2786
 QY 1369 ---KOSIDTIDFANDVLGYKILSEKYKSDLSIKKIYINDKOGENE--KY-----LP 1416
 DB 2787 EKFEELIDITKVKVLACTIONKYLIIHMKIERSTMDIK--IDEKKTIGKIDCKYDVTYTKVA 2844
 QY 1417 FLNIIETLYKTV-----NDKI---DLFY--IHLEAKVLNY 1446
 DB 2845 TCEIDTIDSSVYLKHHVHYSTLSRMOKLIIKYPTEKTHEENFVNPFLKDKVL-Y 2903
 QY 1447 TYEKS--NVE-----VKIKELNYLKTIDOKLADFKKNNNFVIADLSTQYN 1490
 DB 2904 NYKRPINIEHILPGALTIDYDRTKIKO--YILRIPIYV--HKDIHF-----SLEEN 2952
 QY 1491 HNNLTKPLSTGVFENLAKTYVLSNLLDGNLQGLMINSQ-----HOCVKKQCPQNSGCFR 1545
 DB 2953 NSLSLTK--QONIIYGNVAKI-----FIHINOYKEIHGC-----DFGKYS 2993
 QY 1546 HLDERECCGLNKKQBGDKCVENPPTCNENNG 1579
 DB 2994 H-----LFTYSK--KPLPNDDCNVITIG 3015

RESULT 30
 PCT-US93-07261-11
 ; Sequence 11, Application PC/TUS9307261
 ; GENERAL INFORMATION:
 ; TITLE OF INVENTION: P1EMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREO
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John H. C. Blasdale
 ; STREET: One Giralda Farms
 ; CITY: Madison
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07940-1000
 ; COMPUTER READABLE FORM:


```

; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: Malayan Camp
; PCT-US93-07261-16

```

```

Query Match          3.5%; Score 296; DB 5; Length 1663;
Best Local Similarity 20.3%; Pred. No. 8.8e-08;
Matches 308; Conservative 248; Mismatches 601; Indels 362; Gaps 69;

```

```

QY 243 KKKKKTIENTEN--LIEESKTTIDKNNKATKEEKKLYAQVDLSYKQLEEAHLIS 300
DB 12 EKKEKRNALKEKKLEQKNDQAKADLYKKESQD-----SSSEKSLKEKNGEA 62
QY 301 VLEK-RIDTLK-----NENIKELLDKINEIKNPPANGTPTLLDKNKIEHEKEIK 355
DB 63 LKREKKEETLKKLEQKREKEKKNKIKONNDALKKNGDKDKIVKPPSPVEKDK 122
QY 356 ELAKTIKFNIDSLFTDPLELEYTLREKKNK-IDISAKVETKESTEPNEYNGVYPLSYN 414
DB 123 EMLKKEKFKQHKIDYEE---RKEKRRMWILRSLRDKLRLEIOLEKNA----- 170
QY 415 DINNALNELNSFGDLINPDYTKPEPSKNITYTDNERKKEINEIKIEKKIESDKSY 474
DB 171 QLESALINELKERASRRPMVMQKMGKDEVDWMYTKYDDEQAEKNGTKDEELTKDGDT 230
QY 475 ED-----RSKSLNDITKEYEKLTNELIYDSKFNNNIDLTNPEKMMGRYSYVEKLT 525
DB 231 EELVETKFGYGMRENALGEL-DEYEE-----RYEK---KRYLTK----- 264
QY 526 HHHTFASYENSKINLEKTKALK-----YMEDYSLRNIVVEKELKYKILSIEIEFT 560
DB 265 -----EDGEDDLKDVEEKLEETGEGFEKREKPTTIRILVKKR-----RNKEQK----- 305
QY 581 IVENIKKDEQLEFEKKITRKENKPKDEKILEVSD-----YKVOVQVYLLMKXIDELKKTQ 635
DB 306 ---KLKEDK---EKKLIAEPPDDEKKIKLKSDDKVVPV----- 341
QY 636 LILKNVELKHNHIVPSYKQENKQEPYLL-----VLKKEIDKLKVPMPKVESTLINEE 688
DB 342 -----KNKSSFPDKFRAPDKKRTMFYLSLFLPIVPPKD--NELAVCGDSMDSKYNGK 392
QY 689 KKNIKTEGQSDNSEPSTEGEITGQATTKPGQAGSALESDSVQ-----AQA 734
DB 393 KL-----KSTFNPFKRRRNKLKERKMOELHKFKKKNKYKQKLE 431
QY 735 QEQKQKQPPVPPVPPAKQAVTPPPAPVNNKTENVS-----KLDYELKYLEFLTSTICH 769
DB 432 RKRREPPDGPILNTPETIHV---IRPSDLMDKGENKSGAHGPFKYQPTKGLKEY--EESHVSK 487

```

```

QY 790 KYILVSHSTNNE-----KILKQYKITEEESKLSLSCDPLDLFNIGNNIPVYMSFDS 842
DB 488 DYQL-EHEPPTKLPEYKGVNSREYOLDHEPPTKLPEYE-----KGVNSREYOLDNE 538
QY 843 LNNLSQLFMEIYEKEKVCYLKIKONDKIKNLLEKAKKYSTVKTLSSSMQPLSLTPQ 902
DB 539 VRDELPE---YKGVNSREYOL-DNEGPSLKEVDQTELAGKIDITNKPHE--SVDEX 590
QY 903 DKPEVSANDDTSHSTNLNLSLKLFFENILSLGKNKI-----YOEL-----IGQ 945
DB 591 DQTELAGKIDITNKP--HEVSDEYDQ--SELAGKIDITNKHESVDEYDQTELAGKKEVTN 647
QY 946 KSSSENFYEKTLKSDPTFNYSFTNPFYKSKADINSLDESKR---KLEEDINKLKT- 1000
DB 648 KPHENLEE--YNETDLAKGKEVTNKPHEVSDEYDQ--SELAGKIDITNKPHEVSDEYDQTE 704
QY 1001 LQSLFDLNNKYKLKLEPLPKKTKVGYKKNQIKLILKQLESKL---NSLNNPKHVLD 1057
DB 705 LAKGKEVTNKAARENLEEYNETDLAKGK-EVTNKAARENLEEYNETDLAKGKEVTNKAH--E 761
QY 1058 NFSVFENKKEAEIETENTLENTKILKHYKGLVYKNGESSPLTKLSSESIQTEEDNYA 1117
DB 762 NLEEY---NETDLA-----KKEVTNKAHENELEEYNETDLAKGKEVTNKA-----HE 805
QY 1118 SLENFVYLSKLEGLKDNLT---NLEKKSLSYSSGLHILIAELKEVYKKNYTGNSPSEN 1174
DB 806 NLEEYNETDLAKGKEVTNKAARENLEEYNETDLAKG-----KEYV-----NKAREN 850
QY 1175 NTDVNNALSEYKAF-LPEGIDVATVYSESGDPLLEOSQPKPASTHVAGASNTITTSQNY 1233
DB 851 -----LEEYNETDLAKGKEVTNKAARENLEEYNETDLAKGKEVTNKAAREN---LEEYNE 900
QY 1234 DDEVDDVYIPIRGSEBEDDGLGOVYVTEGAVPVSVIDNTLSKIENTEY---EVLVYLKPLA 1290
DB 901 TDLAKGKEVTNKAARENLEEYNET--DLAKGKEVTNKAARENLEEYEDQKKNELQNGKSD 959
QY 1291 GYVRSLK---KQLEN-----NYMTFNVYVKDILNSRENKREPNK-----VL 1329
DB 960 GLKENELKKEELRNKSGDGLKENAELKNEKELRNKSGDGLKENAELKNEKELRNKSGDGLK 1019
QY 1330 ESDLIPIKDL-TSSNIVVADPYKFLKKEKRDKLS-----YVITDSIDTDI 1376
DB 1020 ENAELKNEKELRNKSGGLKENAELKNEKELRNKSGGLKENAELKNEKELQNGKSGELENA 1079
QY 1377 NFANDVL---GYKILSEKKSDDLSTIKYIYNKQENKEYLPLFNNIETLYTVND-KI 1432
DB 1080 ELKNEKELQNGK-----SEGIAKENABOKNKELOKNGSGLEKENAELKNEKELRNKSGDGLK 1134
QY 1433 DLFVILHLEAKVLYTYEKSNEYKIREL-----NYLKT--IQDKLADFK- 1474
DB 1135 NAEELKNEKELRNKSGDGLKENAELKNEKELRNKSGSEGLKENYVTNNDLKNNDIQKDLISNKD 1194
QY 1475 -KNNNFVGADLSTDYNNHNLTKFLSTGVNPPENLAKTVLSN--LDDGNQGLNLNSQHQ 1531
DB 1195 MKNKEILNKDISKDKNNKELLNKLDLSN---EDMKNEKELNLDINKDKLISGNMEQON 1250
QY 1532 CVKQCCP---ONSQCFRHLDEREECKCLNTYKQEGDKVYENPN-PTCNENNGCCADAK 1586
DB 1251 TGLKNTSPSKGQNTGLKNTPNEROQNTGLKNTSPSEGOQNTGLKNTSPSEGOQNTGLKNTPN 1310
QY 1587 CTEEDSG-----SNGRKIT 1600
DB 1311 EROQNTGLKNTSPSEGOQNT 1329

```

```

RESULT 32
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman

```

```

; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572.191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
; US-09-572-191-2

Query Match          3 5%: Score 294; DB 4; Length 1388;
Best Local Similarity 19.4%; Pred. No. 9e-08;
Matches 29; Conservative 240; Mismatches 510; Indels 467; Gaps 70;

QY 93 SVASGASVAGSGNSRRTNPSDSDAKSYADLKHRYNNYLLTIKEL-KYPOLEDLT 151
DB 207 SAARAYVLSGGMRRRVAASMSNRSSRS-----HAV--FTTIESMEKSNELVIR 257
QY 152 NMLTLCDNHGFKYLLIDGEE-----INELL-----YKLNFEYDLRAKLNDVCAN 198
DB 258 TSLINLVD-LAGSEROKDTHEAGMRLEAGNINRSLGQVITALVDGNGKORHYCYR 316
QY 199 DYQCIPIPNLKRAMELDVYKLVGYRRKPLDINIKDNGCKMEDIKKNKKT- 249
DB 317 D-----SKLTF-----LNDSDLS-----GNKTAIIANVHPGS 344
QY 250 -----ENINELIESKKTIDKKNNAKTEBEKKKLYOQYDLSLYNKOLEE----- 294
DB 345 RCFEETISTLNFARAKLIKAKAVNEDTQGNVQOLAQVARKQLQALASGOTPRESEF 404
QY 295 -----AHNLISYLEKRIDILKKNENIKE-LDDKINIKPNPPANSNGTPTLLDKNKI 347
DB 405 LTRDKKNTNMYEYQOAMLEFFKSEOEKSLIEKVTOLED-----LTLKKKEF 452
QY 348 EEHEKEIKELAKTIKFNDSLF-TDPLELEY---YLREKKNNI--DISAKYET-KESTER 400
DB 453 IQSKMIL-----VKFREDDIIRLEKLHKESSRGFLPEBQRLSELNENIOTLAEOTE- 505
QY 401 NEYNGVYTPSYNDINNALNELNSFGDLINPFDYTKEPSKNIYTDNERKAFINEIKKI 460
DB 506 -HHRVAKYAME-----NHSLEENRRLRLLEPVRAQE--MDAQTALEKAFISLSGME 558
QY 461 KIEKKKIESDKKSTEDSKSLNDITKEYEKLINELIYDSKFPNNIDLTNFEKMGKRYSK 520
DB 559 KSDKNQOGFSFKAQKEPCLFAN--TEKLKQOLQIOTELANSKOYEFEKELTKRR--- 612
QY 521 VEKLTNHTFASYSNKHNEKL---TKALKYMEDYSLRNIVE-----KELKYYK 568
DB 613 --QLELESELOKANINLENLEATKACKROEVSQLNKHAETLKITTPRAYOYKIL 670
QY 569 NLISKIENIETLEVENIKKDEOLFEEKITDENKPD--EKILEV--SDIVKVOYOKVL 624
DB 671 RPYVKLSPEMGSF--GSLYTQSSSLDNDILNEPVPREKNEQAFELISELTYOQOMSL 729
QY 625 MNKIDELKKTOLLKNEVLEKHNHVPNSYKQENKQEPYLLIVLKEIDKLKVPKPYESL 684
DB 730 QAKLDE-----EEHKN-----LKLQOHDKLEHNSHTOMEL 760
QY 685 INEKKNIKTGSDNSEPSTEGEITGOATTKRQOQAGSALGSGSVQAQAOQOQADQRPV 744
DB 761 FSSERID-----WTKQOEELLSQL--VYLEKQOE----- 788
QY 745 PVPVPEAKAQVTPPAPVNNKTEVNSKLDYLEKLEFPLNTSYICHKYLIVSHSTNEXIL 804
DB 789 -----TOTKN-----DPLKSF--VHDLRVVLHSA----- 810
QY 805 KQYKITEESKSLSCDPLDLFLNIONNIPVYGMFDSLNSLSQLEMEIYEKEKAVCNLY 864
DB 811 -----DKELSS-----VKLEYSSF--KTNQKEFKKLSERHMHVQL- 844

```

```

QY 865 KLKNDKIKNLLEKAKVSTVKTLSSSSMQPLSTTPQDKREVSAN-----DDTSHTN 918
DB 845 -----QDNLRLNEKELLESKACLODYDNLQELIMKFEIDLSRLNLONEKKEMLAKSD 898
QY 919 LNSLKLFEENILSGKNKIYOEILIGOKSENEFEKILKSDTFYNESEFTNFKSKADDI 978
DB 899 LNNIMELLE-----AEKENNNLSLOFEEDKENSKEILKYLEAVROEKQKTACE----- 950
QY 979 NSLDESKRRKKLEEDINKKKTLOLSPDLNKKYKLEKLELPDK-KTYVGKTKMOIKLITL 1037
DB 951 --OQMAKVOKLEESLATEKVIS-----SLEKSDSPDKVYVADLMNDIOE--- 993
QY 1038 LKEOLESKLNINPNKHVLONFVFN-----KKKEAETAEENTLENTKIL 1084
DB 994 LRSSVCKETETIDYLQELKINCKYSALVDRRESRYLIKQEVYDILDKETLR----- 1048
QY 1085 LKHYGKLVKYYNGESSPLKTLSEE--SIQEDNYASLENPKVL--SKLEGKLDNML 1138
DB 1049 -----LRLISEDIERDMLCEDLAHATEQLMNLTASKKSHGLQSAQE 1091
QY 1139 EKKKLSYLSGLHMLIAELKVIKKNKTGNSPSENNTDVNNALESTYKFLPEGTDAVY 1198
DB 1092 ELTKKREALIOELQKLNQKKEVEQK-----NEYNFKMQO-LEH 1130
QY 1199 VSESGDTELOSOPKRA--STHYGASNTITTSQNVQDEVQVYIYIFGESEDEYDYL 1256
DB 1131 VMSDAE--DPQSRKTPPHFQTH-----LAKLTLOEOEIED- 1165
QY 1257 GQVYTGAVTPSVIDNLSKINENEVLYLKLPLAGVYRSKLQENNVMTFNVAVKILN 1316
DB 1166 -----GRASKTS-LEHUTTKLNEDREVKNAIL-----RMEQL----- 1198
QY 1317 SRFKRENFKNVLSDLPIPYDLTSSNIVVADPYKFLKERRKDXFLSYN-----IKDS 1371
DB 1199 --REMEMLR--LES-----QOLIEKMMLQOQ--LDDIKROKENSQNNHPDQOLKNE 1245
QY 1372 IDTDINFAVDVLYGYYKILSE--KYSKSDLSIKKIYNDQGENEYKLPNNIETL-YKTV 1428
DB 1246 QEBST--KERLASKIYEEMLKAKADEEVSQALYNKEMECINMTDEVERTOTLESFAF 1302
QY 1429 NDKIDLFYHLEAVLNTYKESNVEYKIKELNLYKTIQDKLADFPKKNNFVGLADLSTD 1488
DB 1303 QEKEDL-----RSKLEEMYEER--ERTSQEMEMLRKQVECLAE--ENGKLVG----- 1345
QY 1489 YNHNMLTK 1497
DB 1346 --HONLHOK 1352

```

```

RESULT 33
US-09-723-262-2
; Sequence 2, Application US/09723262
; Patent No. 6379912
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723.262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572.191
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
; US-09-723-262-2

```



```

Db 317 D-----SKLTF-----LIRDSLG-----GNAKTAIIANVHPGS 344
QY 250 -----ENINELIEESKKTIDKNNKNAFKEEKKKLVQAOYDLSTYNKOLE----- 294
Db 345 KCFGETLSTLWPAQAKILKNAKAVNEDTQGNVSOLQAEVRLKQLELASSGOTPPPSF 404
QY 295 -----AHNLISVLEKRIIDLTKKNENIKE-LUDKINEIKNPPANGSPTMLTDKNKI 347
Db 405 LTRDKKKNVMEYQEAALFFKKSQOEKSLIEKVTQLED-----LTLKKKEF 452
QY 348 EHEKEIKEIKTIKENIDSLF-TDPLELEY--YLREKKNKI--DISAKVET-KESTEP 400
Db 453 IQSNKMI-----VKFREDOJIRLEKIKHESGGLPEEORRLSELNEIOTTLQEOLE- 505
QY 401 NEYPCGVTPPLSYNDINNALNELNSFGDLINPDYTKEPSKNIYDNRKKRFINIKKI 460
Db 506 -HHRPVAKYAME--NHLSEBNRLRLLEPVKRAOE--MDAOTIALEKAFESTISGME 558
QY 461 KIEKKIESDKSYEDRSKSLNDITKEYEKLNLNEYDSKFNNNIDLTNFKMMGRKYSYK 520
Db 559 KSDKNQOGFSPKAQKEPCLFAN--TEKLKAQLOIOTELNNSKQOEYEFKELTRKR----- 612
QY 521 VEKLTNHTFASYSKSNLEKL--TKALKYMEDYSLRNIVF-----KELATYK 568
Db 613 --OLELESELQSLQANLLENLLEATKACKROEVSQLNKIHAEYLKITTPTKAYQLHS 670
QY 569 NLISKIENEIEITVENIKKDEEOLFEKKITKDEKNPD--EKILEY--SDIYKVQOXYKL 624
Db 671 RPKVKLSEMGSF-GSLTTONSSIINDILNEPRPENNEQFEELISELRTVQOMSL 729
QY 625 MNKIDELKKTOLILKNVELKHNHIVNSYKQENKOEPRYLILVLEKIDKLKVFMPKVS 684
Db 730 QAKLDE-----BEHKN-----LKLQOHVXKLEHSHQOMEL 760
QY 685 INEKKNIKTGQSDNSEPSTEGELTGOATTKPGQOAGSALBGDSVQAOBOKQOAPRV 744
Db 761 FSSERID-----WTQKOEELLISOL--NVLKQLOE----- 788
QY 745 PVPVPEAKQVTPPAPVNNKTEVNSKLDYLEKLEFINTSYICKYILVSHSTWNEKIL 804
Db 789 -----TOTKN-----DPLKSE--VHDLKRVYLHSA----- 810
QY 805 KQYKITEESKLSGCDPLDLFLFNIONNIPWYSMFDLSLNSLSOLFMEIYEKEWCMILY 864
Db 811 -----DKELSS-----VKLEYSSF--KINOKEFEFKLSRHHHVLQ- 844
QY 865 KLKNDKIKNLLEAKKAVSTVKTLSSSQPPLSLTPDQKPEVSAN-----DTSHTSN 918
Db 845 -----OLDNLRLENEKLLLESKACLODSYDNLQEIIMKFEIDQSLRNLONFKNENETLKSD 898
QY 919 LNSLSKLEFENLISGKNKNITVOELLIGOKSSSENFYEKILKSDDTFYNEFTNFVYSKADI 978
Db 899 LNNIMELLE-----AEKENNNKLSLOFEEDKENSSEKLEILKLEAVROEKOKETAKE----- 950
QY 979 NSLNDSEKRRKLEEDINKLKTLOLSPDLNKKYKILKLELFDK--KTYGKYMJOIKKLT 1037
Db 951 ---QQMAKVQKLESLSLATEKVIS-----SLEKSDSDKKVADLMOQOE--- 993
QY 1038 LKEOLESKLSLNNPKHVLQNFVFFN-----KKKEAEIAETENTLENTKIL 1084
Db 994 LRSVYCEKTEFTIDTLKQELKIDCKYNKYNALVDRRESRVLLIKQOEVDIDLTKETLR----- 1048
QY 1085 LKHKGLVKKVYNGESSPKLTSEE--SIQEDNVAJSLENKVL--SKLSEGLKLDNML 1138
Db 1049 -----LRLISEDERDMLCEDLAHAEQULNMLLEAKKSHGLQSAOE 1091
QY 1139 EKKRLSYLSGLIHLIAELKEVIRKKNYTGNSPSENNTDVNNALLESYKPLPEGTDAVTV 1198
Db 1092 ELTKKALIOELQHLKLNQKKEVEBOKK-----NEYNFKMQO-LEH----- 1130
QY 1199 VSESGDITLQESQPKRA--STHYGAESNTITTSQNDVDEVDAVTVIPFESSEEDYDL 1256
Db 1131 VMSAAE--DPOSPKTPPHQOTH-----LAKLTLETOEOEIED- 1165

```

```

QY 1257 GOVVTGEAVTPSVIDNILSKIEENEYEVLYLKLPLAGVRSKLKQENNVNWFENVAKDILN 1316
Db 1166 -----GRASKTS-LEHVTATKLEDEKVEKNAEL-----RKQEL----- 1198
QY 1317 SRFNKNREKKNVLESDDLPRYDOLTSNNVVDKPYFKLEKRRKDFLSSYV-----IKDS 1371
Db 1199 ---REMENLKR--LES-----QOLIEKNMLQOQ---LDDIKRQKENSQDHPDNOQLKNE 1245
QY 1372 IDTDINFANDVLGYKTIIE--KKSQSDLSIKKIYNDKQGEKEKYLPLNINIELT-YKTV 1428
Db 1246 QEESI---KERLAKSIVIEEMLMKRADELEVOSALYKNEEMCLBMTDDEVERTQLESKAF 1302
QY 1429 NDKIDLFIYHLEAKVLANTYKESNVYKIKELNYLTQIDOKLADPKKNNPFVGIADLSTD 1488
Db 1303 QEKEDL-----RSKLEMYEER--ERTSQEMMLRKQVECLAE--ENGKLVG----- 1345
QY 1489 YNHNMLTK 1497
Db 1346 --HQLHOK 1352

RESULT 35
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

Query Match 3 5%; Score 293 5; DB 4; Length 1010;
Best Local Similarity 20.0%; Pred. No. 6.5e-08;
Matches 240; Conservative 201; Mismatches 417; Indels 341; Gaps 56;

QY 226 KPLNDIKDNVGMEDYIKNNKKTIE---NINELIEESKKT-----IDKN 266
Db 3 KPLHIVMENFGPF-----IKETIDFEQVEVDOLFLLGSKTGSGKTMIFDAIVAYLYGMA 56
QY 267 KNATKEE-----EKKRLVQAOYDLSIYKQOLEEAHNLISVLEKRIIDLTKKNENIKEL 318
Db 57 STKTRKEGDLRSHFADGSPMSVIVQFVNNQTFKIHREAPFIEGNTTKTOAKLNIYEL 116
QY 319 LDKINEIKNPPRANGNTPNLL-----DKNKKIEHEKEIKIATYI 361
Db 117 VDNQFELRE--SKVQGNQPIVQLLGVAEQRQLEFILPDQGEKFFLOSNKDKOSILFTL 175
QY 362 KENIDSLFTDPLELEYULREKKNKNDISAKVETKESTEPNEVPNGVTVPLSYNDINNALN 421
Db 176 -FNSFRPD-----ELRHLLVENVKQEV--QIENR-----YTQENLWLN 211
QY 422 ELNSFGDLNPFYDKPSKNIYDNERKKF--INEIKKIKIEKKKIESDKSYEDRSK 479
Db 212 DIDFENN--DELAUYKE--LESSQDCKMIKFPQFNQYCKI-----LKSFE--- 255
QY 480 SLNDITKEYEKLNIYDSKFNNNIDLTNFKMMGRKYSYVVEKLTHTNHTFASYSKSN 539
Db 256 AKNKTKLELDL-----NHKRYVNEISENTRKL-----KAEKI---KPDOLKKEBOY 300
QY 540 LEKTLKALKYMEDYSLRNIVVEKELKYKNL--ISKIENEIEITVENIKKDEEOL----- 592

```

```

Db 301 IDKLKOEKLMIOESKVL-----ITFYTRLOSLKKKDEKDELAVSLHROSKLNETFNHNEIK 353
OY 593 -PEKKITTKENKPE-----KILEVSDIYKVQVKVILNMK-----IDELK-----TOL 636
Db 354 GKOKOLEHSTRNETTOFNQYLEKNOVFPNOJDKITSSYQOKPVATEELKRLYSEYNDL 413
OY 637 ILKNVELKNIHVPNSYKOEKOEPEYLLIV-LKEIDKLVKEMPKVESLINEKKNIKTE 695
Db 414 ITRKEEL-----TKEMNNKMKDFALLIEHYEELIKLK-----KIIDSEBROKDE 458
OY 696 GOSDNSEPTSEGTGOATTKPGQAGSALGDSVOAQOEOQAPPVPPVPEAKAQY 755
Db 459 KLEFDKQ-----LDKSSYLSKLKEKKEQOINEISSITINIDATL 496
OY 756 PPPPAPVNNKTENVSKLDYLEKLEFELNLSYICHKYI--LVSHSTNNKELKKYKIKTE 813
Db 497 ----IDLNDKKQVNE-----IKSAMISIGDTCPIGNEIHSIGEHIIDESAQKNNKIKLE 549
OY 814 ESKLSSCDPLDLLENIONNIPVYMSFDSLNNSLSQLEMEIYERKVCNLK-----L 866
Db 550 SKKV-----KIRDEIKIETRIEELNHNENELNFEKQEKDISELQKQNLNHLNOL 599
OY 867 KD-----NDKIKNLDEAKVSTSVATLSSSSQPLSL-----TPQOKPEYSANDOTSHS- 916
Db 600 KEOQOSINKLVENFEKOEKEIVNKHOFD-----LDLSRKNQKREKLEIOINDFERHSQ 653
OY 917 -----TNLNSLTKLFENIISLKKNNKIYOELLG 944
Db 654 FSSVNDPEFYIYSHAKKOVETYEYENKTKDKLNLKNNK-----IEMNQKILTEMLT- 707
OY 945 QKSS-----NFEYKILKDSDFPYNESFTNPKSKADINDSLNDESKRRKLEEDINKLKTKIO 1002
Db 708 QJSKEINNNLELKKEMQOLGFESEYDQ-VKSAA-DLSAQKDE-----IEREINIYNNK-- 758
OY 1003 LSRDLNKKYKLEKLEEDKKKYKMOIKKYLTLKEOLESKLNSLNKNNKHLQNSVF 1062
Db 759 -----YQSEIEINRL--KELYGKKLNLDELRQSTIEKTNLKLDET- 799
OY 1063 FNKKRAEALTEPNTLENTKILKHKYKGLVYKNGESSPLK-----TLESSEIOTEDNVASTL 1119
Db 800 -----SQATISYKIDNNSNKNFNKIKNIIOIIDDELKVKQKEIFELSELLAGNDKRLTL 853
OY 1120 ENPKVYSLKLEGLKDLNLKELKKLSTLSSGLHLHLAEKLVY-----KNN 1164
Db 854 ENYVLLIYVLE-KIIFQAN--ORLSFMSGNRYOLIR--RETIISLGLSEIDVDFEHSNK 907
OY 1165 NYGNSPSENNTVNNALLESYKKEFLPEGTQVATVY--SESGDPLEQSQPKKPASTHGAE 1223
Db 908 SRIISLSG-----ETFOASLALAGLSEVVOQESGCTILD-----SMFIDEG 951
OY 1224 SNTITTSQNVDEVDVITVPIFGESEEDYDDLGQVVTGEAVTPSYDNLISKIENEYE 1282
Db 952 FGTL-DQETLETAIDLTLINKSSGR-----WGIISHVSELKORRPLLEVTLSNQYE 1002

```

```

RESULT 36
US-08-446-855A-2
: Sequence 2, Application US/08446855A
: Patent No. 5849573
:
: GENERAL INFORMATION:
: APPLICANT: Stewart, Thomas S
: APPLICANT: Flores, Maria V
: APPLICANT: O'Sullivan, William J
: TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
: TITLE OF INVENTION: phosphate synthetase II
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon & Vanderhye PC
: STREET: 1100 No. 5849573th Glebe Road, 8th Floor
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA

```

```

: ZIP: 22201-4714
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/446,855A
: FILING DATE: 06-Jul-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Mitchard, Leonard C
: REGISTRATION NUMBER: 29,009
: REFERENCE/DOCKET NUMBER: 47-80
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-816-4000
: TELEFAX: 703-816-4100
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2391 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-446-855A-2

Query Match      3.5%; Score 293.5; DB 2; Length 2391;
Best Local Similarity 19.1%; Pred. No. 1.9e-07;
Matches 323; Conservative 252; Mismatches 543; Indels 577; Gaps 83;

OY 171 YEEIN-----ELAKLNFYDLRLAKINDVACANDYCOIPPNKIKRANEDLVKKIYF 222
Db 232 YKEINLEDPGNIDTLKYVCHNFIKRVI--KLN-----ITVNYK-----NKEEF 272
OY 223 GYRKPLDNIKDNVGMKEDYKRRKKKTIEININELIEESKTIIDKRNKATKEEKKLYQAO 282
Db 273 NY---TNEKTTNDSMEDHNEINGISNPNNC--PSISFDS-----ESKNV---- 316
OY 283 YDLISYNNKOLEAHNLISYERKIDPLKKNENIKELDKINEIKNPPANSGMTPTLTD 342
Db 317 ----INHTLRDKNMNTITSSEYELKDL-HNCFNSNSDK-----NDSEF 355
OY 343 KKKIEHEKEIKETIKTF-----NIDSLFTDLELEYLREKNNKIDISAKYETSESTE 399
Db 356 KLYGIEYDYLIDLEENASFHYNNVD-----EYGYVNNKNTNLSNKKIEQNNNE 408
OY 400 PNEYGVVYPLSYNDINNALNEL-----NSFGDLNPFDTYTESKNITD--NEK 450
Db 409 NNR-----NNKNNNNEVDYIKKDEDNNVNSKVYSQYNNNAQNNHEHTTEENLN 457
OY 451 KETNEIKETIK-----IEKKIESDKSYEDRSKSLNDITREYKLEINEIYDSKFNNN 503
Db 458 DYSTYIRKKKKNEEFLNVLVKKRVDDKEK-----IIVYDCGIKNS 498
OY 504 IDLTNEKMMGKRYSYKVELTTHNTPASE-----NSKNNLEKTLKALKYMEDYSLRN 558
Db 499 IIKNLIRHGMDPLTYIIVPYYNENHIDDAVLLNSNGEDPKCOPLINKLNDKSLTKNK 558
OY 559 VV-----EKKLKYKNLSIKIENEIETLVENIKRDEO-----LFEK 596
Db 559 IIFGICLGNOLGISLSCDTYKMKKYGNR---GYNQVYQIVDNYCITYTSNHHQCLKKKS 615
OY 597 ITRDE-----NKPEKILEVSD---IVKVOVQVLLNKKIDELKKTQLILKNVELKH 645
Db 616 ILKRELAISYINANDKSTIGISHKNGRFYSQFHP-----EGNNGEDTSTFLKKNFL-- 669
OY 646 NIHVPSYKOEKOEPEYLLIVLKKET----- 671
Db 670 --DIFKKKQRYRELYGNIITYIKKKVLLSGGLGICIGQAGEPDYSGTOAISKLECGIYV 727
OY 672 -----DKIKVEMPK-----VESLINEKKNN--IKTEGOSDNSEPTSEGE 708
Db 728 ILVNPNIATVOTSKGLADKY-YFLPVNCEFEVEKIKKEKPDFILCTFG----- 774

```



```

QY 709 ITGQATTPG-----QAQSALEGDSVOAQ-----AOEQOAPVPVPVPE 750
  |||
Db 775 --GOTALNCALMDOKKVLKKNQCCTSIESTIRITENRLEFAKLEINERI-APRGS 831
QY 751 AKAOVPPAPVNNKTENSVKLDY--LEKLYEF--LNTSYI-----CHKYILVSH 796
  |||
Db 832 AK-----NNOAIDIANIKGPIVLRFTFSIGLNSPFINNEEELIECKNIFL--- 880
QY 797 STAME-----KILK-----QKITKEES-----KLSSCPDLLEFNIONNIPWYSMED 841
  |||
Db 881 QTDNEIFIDSLQGWKEIEYELKDNKNNCIALCMMENIDPL-----GIHTGDSIVAVASQ 936
QY 842 SLNLSLSQLFMEIYEKEMV-----CNL--YKLKDNKIKILBEAKKVSVTLSLSSS 893
  |||
Db 937 TLSVYEYKREILAKVITHLNIIEGCIQIOPGIMPOTGEICIIENARLSKSLASKAT 996
QY 894 MQLP-----SLTPQDK-----PEVSAANDTSHSTNLN 921
  |||
Db 997 GYPLAYISAKIALGYDLISLKNSTIKTKTACFEPSLDYITTKIPRMDLKNREFASNTMNS 1056
QY 922 SKLFENILSLGKKNTYOELIGOKS-----SENFY-----EKIL-----K 957
  |||
Db 1057 SSKSVGEVMSIGRP--FEESI--OKSLRCIDNYLGFENYCIDWDKCKIIEELKNPSPK 1112
QY 958 DSDTFYNESFTNPFYKSKAD-----INSLNDSKRRKLEE--DINKLKK 999
  |||
Db 1113 RIDAIIHQAFLHMPMDKIHETLIDYFPLHKEFYNIYINLQNKLYLQSLQSFNDLKYFK 1172
QY 1000 -----TLQSFEDLYN-----KYKLELRLDKKTKYKQYKMOI 1032
  |||
Db 1173 HGFSQKQIAHYLSFNTSDNNNNNNNISCRTENDVMKRYRKL--GLRPHIKVIDTSLAEF 1231
QY 1033 KKLK-----LKEQLESKLSLNPKNHVLQNSVFENKK-----EAEI 1071
  |||
Db 1232 PALTNLYLYTYOGQHDVPLNMMKRIKICFLNNKRNA-----NKKYHVKNHLYNEV 1283
QY 1072 AETENT-----LENTKILKHYKGLVYKNGESSPLKTJSESIQEDN 1115
  |||
Db 1284 VDDKQDLOLHKNNNNNNNMANGVENEKCKLNKESG--YNNSSNCINT--NNINIEBN 1336
QY 1116 ----YASLENFKY--LSKLEKLDKNLMEKKKLSYLS--SGLHHLIAELKEVI----- 1161
  |||
Db 1337 ICHDISINKNIKVTINSSNSISNENVE--TNLCVSEBAGSHIYKEEKESIGSDPTNI 1395
QY 1162 ----KKKNTGNSPSEN--NTDVNNALESYKKFLPGCTDAIYVSSGSDTLEQSOQPK 1214
  |||
Db 1396 LSAONSNNNNSCENNENKKNANVDVNLENDTK--REDINTTIVFME--GQNSVINNNKE 1452
QY 1215 PASTHVAESNTITTSQNVDEVDVILVPLFGSEEDYDOLGOVVGAVTPSVINIL 1274
  |||
Db 1453 NSSLSLKDEDEDIYVNLKKNENNSVI-----NNVDCRKKMOQK--NINDECK 1499
QY 1275 SKIENEYEVLYLPLAGVYSLKQLENNVM-----TFNVNVKDLNLSR-----F 1319
  |||
Db 1500 TYKKNNKRYKDM-----GLNNNIYDELNGSHSTNDHLIDNFNTSDEIGNKNMMDYL 1553
QY 1320 NKNENFNVLESDLIPLKDLTSSNYVVKDPRKFLNKEKROKFLSSYTIKDSITDIN-- 1377
  |||
Db 1554 SKESKISLN-----KNPGNSYYVDSVYNNMEYKINKKKELEIDNENLNDENYNNNNVM 1604
QY 1378 -----FANDVIGYKILISEKYSKSDLSIKKY--INDKOGENEKYLPFLNNIELY 1425
  |||
Db 1605 CSNTNNAASAVNGKDRDNLENDCIEKNMDHTYKHYNRLNRRSTNERMMALVNN----- 1659
QY 1426 KYVNDKIDLEFVILHLEAKVLYNT--YEKSNEYEVKIKELNLYLKTIDOKLADFKNNNFVIA 1483
  |||
Db 1660 -----EKESNHEKGHRNGLNKKKKEKMEKN--KGNKMD--KKNYHYVN-H 1701
QY 1484 DLSTDYHNNLLTKF 1498
  |||
Db 1702 KRNNEYNSNIESKF 1716

```

```

RESULT 37
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: A093/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

Query Match
3.5%: Score 293.5; DB 4; Length 2391;
Best Local Similarity 19.1%: Pred. No. 1.9e-07;
Matches 323; Conservative 252; Mismatches 543; Indels 577; Gaps 83;

QY 171 YEEIN-----ELLYKNFYFDLRAKLNDVACANDYCOIPENLKIRANELDVLKLYE 222
  |||
Db 232 YKEINLPDPCNIDILKVCVNHFIYVI--KLNN-----ITYNYK-----NKEEF 272
QY 223 GYRRPLNINQNMGMKMDYIKKNTJENINELIESKTTIDKKNNTKKEBKLLQAO 282
  |||
Db 273 NY--TTEMITNDSSMEDHNEINGSISNFENC--PISSEFDSK-----ESKNV----- 316
QY 283 YDLSTYKQLEEAHNLISYLEKRIIDLTKKNENIKELDKIENIKNPPANGNTPTLLD 342
  |||
Db 317 ----INHTLRDKNNLITTSSEIYKLD--HNCNFSNSSDK-----NDSFF 355
QY 343 KKKIEHEKEIKETIAKTF--NIDSLFTDPLELEYLREKKNIDISAKVETKESTE 399
  |||
Db 356 KLYGICEYDKYLDLEENASPHYNNVD-----EYGYDVNKNNTILSNKKIQNNNNE 408
QY 400 PNEXPNGVTYPLSYNDINNALNEL-----NSFGDLINFDYTKEPSKIYTD--NERK 450
  |||
Db 409 NKK-----NNKNNNNNEVDYIKKDEDNVNSKVFYSOYNNAONNEHTPEMLN 457
QY 451 KFINEIKERK-----IEKKKIESDKSYEDRSKSLNDITKEYEKLMEIYDSKFRNN 503
  |||
Db 458 DYSYTIKKKKKNEEFMLVNNKRYVDHKE-----IYIVDCGATKS 498
QY 504 IDLTNFKMGKRYSYVEKLTTHNTFASYE-----NSKHLEKLTALYMEYSLRNT 558
  |||
Db 499 IIKMLIRHGMDLPLYIIVPYYNFNHIDYDALLSNGPGDKCDPLIKMLKSLTRKK 558
QY 559 VV-----EKEKLYYKNLSIKIENELJETLVEIKKQDEO-----LFEKK 596
  |||
Db 559 IIFGICIGNOLLGISLSCDPTYKMYGNR--GVNQPVIOLVDNCTITSQNHGICLKKS 615
QY 597 ITKDE-----NKPDEKILEVSD-----YKVQYQKVLNKKIDELAKTOLILANVELKH 645
  |||
Db 616 ILKRELALSYINANDSIEGISHKNGRFYSVQHP--EGNNGPBDTSPLFRKNFL-- 669
QY 646 NIHVPSYKQENKOEPLYLLVKKET----- 671
  |||
Db 670 --DIFNKKQYREYLYGNYIYIKKKVLLSGGLGICGAGEFDYSQTQAIKSLKECGIYV 727

```

QY 672 -----DKLVPMK-----VESLINEKKN-ITKESGDSNESPTEGE 708
DB 728 ILVNPNIATVQTSKGLADKV-YFLPVNCEPVEIKKEKPDFILCTFG----- 774
QY 709 INGQATTKG-----QOAGSALGSDSVQAO-----AQOQAOQAPVPVPVPE 750
DB 775 --GQALNCALMDOKKVLKNNCCGLGSLIESIRITENRFLFAELKLEINERI-APYGS 831
QY 751 AKAQVPTPPAPVNNKTEVNSKLDY--LEKLYE--LNTSYI-----CHKYLVSH 796
DB 832 AK-----NVAQDAIDANKIGYPLVVRTFTSLGSLNSFINNEBELIEKCKIFL--- 880
QY 797 STWNE-----KILK-----QYKITEES-----KLSGCDPLDLFNIONNIPVMSWFD 841
DB 881 QJDNELFIDKSLQGWKEIYELRDKNNCAICNMENIDPL-----GIHTGDSIVVAPSQ 936
QY 842 SLNNSLSQLEMEYEKEW-----CNL-YKLKNDKIKNLEAKKYSTSVKTLSSSS 893
DB 937 TISNTEYFYFRELAKVIITHLNIIGECNIOFGINPOTGEYCIEVVARLSRSALASKAT 996
QY 894 MOPL-----SLTPQDK-----PEVSANDDTSHSTNLNN 921
DB 997 GYPLATYISAKIALGYDILSLKNSITFKTACFEPESLDYITTKIPRMDLKKFEPASMTMS 1056
QY 922 SLKLEENILSLGKNKNITVOLLGOKS-----SENFY-----EKIL-----K 957
DB 1057 SMKSVEGVESIGRT---FEEESI-QKSLRCIDNLYGFSNTYCIDMDDEKRIIEELKNPSRK 1112
QY 958 DSDTFNESEFTNVKSKAD-----INSLNDESKRKLEE---DINKLKK 999
DB 1113 RIDAIHOAHLMPMDKHELTHIDYWFIAKRYNIYNLONKLTLYLEQSLFMDLKYFFK 1172
QY 1000 -----TLQSLFDLYN-----KYKLELRLFDKRTTKYKYMOT 1032
DB 1173 HGFSPDKQIAHYLSFTSDNNNNNNNNISCRVTENDVMKYREKL-GLEPHIKYIDLTSAEF 1231
QY 1033 KKLTL-----LLKQLESKLSLNNPKHYLONFSYFEPKRR-----EAEI 1071
DB 1232 PALTYNLYLYTQGOEHVPLPLMKRRKICTLNNKRNA-----NKKVHYVKNHLYNEV 1283
QY 1072 ATEENT-----LENTKILKHYKGLVYNGESPPLKTLSESIOTEDN 1115
DB 1284 VDDKDTQLHKEKNNNNNNMNSGVEKCKLAKESYG-----YNNSSNCINT---NNINIEKN 1336
QY 1116 ---YASLENFVY-LSKLEGLKLDLNLKELKLSYLS--SGHLIAELKEVI----- 1161
DB 1337 ICHDISINKIKVITINNSNISNENVE-TNLNCYSEBAGSHIYIGKEKSGSDOTNI 1395
QY 1162 ---KNKRYTGSNSEN--NPDVNNALLESYKFLPEGTVDATVSESQSDTLEQSOPRK 1214
DB 1396 LSAONSNNNFSCNENMNKANVDVNVLENDTK--REDINTTTFPME--QONSVINNNKE 1452
QY 1215 PASTHGAESNTITTSQAVNDEVDVYIPIFGESEEDYDDGOVVTGAEVAPSVINDLI 1274
DB 1453 NSSLKGBDEEDIVMNLKKNNTNSYI-----NNDCKRCKMDGK---NINDECK 1499
QY 1275 SKIENEYEVLYLPLAGVYRSLEKOLENNVM-----TFNVAVKIDILNS----F 1319
DB 1500 TYKKNKXKDM-----GLANNITVDELSTNGTSHSTNDHLYLDNNTSDEDELGNKKNDMYL 1553
QY 1320 NKRENFKNVLESDLLPYKULTSSNYVVKPYKFLNKEKDKLSSYNYIKDSIDTIDIN-- 1377
DB 1554 SKKESISN-----KNPGNSYUYVDSVYNNNEYKINKMKELIDENINDEYNNVNMN 1604
QY 1378 -----FANDVLGYIILSEKYSKSDLSIKY---INDOGENEKLPLPLNINLETLY 1425
DB 1605 CSNYNNASAFVNGKORNDNLENDCELEKNDHTYKHYNRNNRSTERMLAVNN----- 1659
QY 1426 KTVNDKIDFVILHLEAKVLYNT--YEKSNVEYKIKELANTLKTJODLADCFKNNNVGIA 1483
DB 1660 -----EKESNHEKGHRNGLNKKKNEKNEKN--KGNKPD--KKNHYIYVN-H 1701
QY 1484 DLSTDYNNHNNLLTKF 1498

DB 1702 KRNEYNSNNIESKF 1716
RESULT 38
US-08-755-587-44
Sequence 44, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-755-587-44
Query Match 3.5%; Score 292.5; DB 3; Length 3418;
Best Local Similarity 18.6%; Pred. No. 3.3e-07;
Matches 382; Conservative 288; Mismatches 717; Indels 663; Gaps 92;
QY 7 LGSFLP-----FIINTQCVTH--ESYQELVKLEALEDAVLGTGSLFQKEKMYLNECTSGT 60
DB 160 VCGSLFHTPKFYKGRQTPKHISSELSGAEDVDPDMSWSSSLATPPTLISSTVLIYRNEASET 219
QY 61 AVTSTPGSKGSVASGSGGSVASGGSVAS--GGSVASGGSVASG--SGNSRRTNPS 114
DB 220 VEPHDTTAVKVSFNSHDESLKNDREFIASVIDSENTNOREASHGFGKTSGNSFKN-- 277
QY 115 DMSSDASDAKYADLKHRYNNYLLTIKELKYPQLFDLTNH-MULTIDNHGPFYTL-IDGY 171
DB 278 -SCDKHIGSMPIV-----LEDEVYETVVDTSSEDSFSLCSKCRTNLQAVRPS 326
QY 172 EEINLELYKINFYFDLLRAKLVNVCANDYCOI-----PMLKIRANLADVLKLVFGYRKP 227
DB 327 KTRKKRIFHPAN-----ADECEKSKQYKKEKSFSEVPNPDTDLDSNV-AHQRP 375
QY 228 LDNIKDNVGMEDYIKKNNKTIEINELIESKKTIDKNNKNAITKEEKKLYQAOYDLSI 287

Db 376 FEESGSDKISK-----EYVPSLACEWSQÜLTLSGLNAQOEKIPLLHISSQDNI 423
Qy 288 YNKÖLEEHANLISYLEKRIIDLTKKNEINIKELLDKINET-KNPPRANGTPTMTLDKNNK 346
Db 424 SEKDLDPTEN-----KRRKDFLTSEN--SLPRISSLPKSEKPLNEETVANK-RDEEQH 473
Qy 347 IEEH-----EKEKEIAKTKEFIDSLF-----TDPLELEY 378
Db 474 LESHTDCLLAKQAIISGTSVPAASSFOGIKTSIFRIRESKETETFNASFCHMTDP-NFKKE 532
Qy 379 LREKKNIDISAKVETKE-STEPNEYPNGVYPLSYNDINNALNELNSEGDLINPDYRK 437
Db 533 TEASRSGEIHVYCSQKEDSLCPNLIDNG-SMPATTTONVALKKAAGLSTL----- 583
Qy 438 EPSKNIYDNEKKKFINIEKEKIKIEKKIESDKSYEDRSKSLNDITKEYEKLNETYD 497
Db 584 -----KKTJNKFIYAIHDETFYKGRKKIPKDOKS-----ELIN-CS 617
Qy 498 SKFNNDIDLTFNEKMMGRYSYKVEKLTHNTEFASYEN-----SKNLEKLTFA 546
Db 618 AOFENNA-----FEAPL-----TFANADSGLHSSVRCSCQNDSEPTLS 658
Qy 547 LKYMEDYSILR-----NIVEKEKLYYKNLISKIENEIETLVENIKKDEOLFEEK 596
Db 659 LTSFSGTILRKCSRNETCSNMNTVISQDLDYKCAKCNK-----EKLQLFITPEADSLSCQ 713
Qy 597 ITKDNKAKDEKILEYSDIVK-----VOYQAVLANKIDELAKTQILKLNVELKINI 647
Db 714 EGQCNDKRSK--KVSDIEEVLAAACHPHVOHSKEYESDTPOQOSQSL----- 760
Qy 648 HVPNSYKOENKOEYLLIVLKEIDKLKVFMPKVESLINEEKNKITEGOSDNSEPTBG 707
Db 761 -----YDHENAST-----LILTPRSKVLNLY-----MISRGESYKMSDKLKGNYTESDY 807
Qy 708 EITGQATTPEQOQAGSALEGSVQAQOEQOQAPVPVPEAKAQAFTPPAPVY-NKT 766
Db 808 ELTKRIPMEKNO-----DYCALINENYKN-----VELLPREKYMRVASFPRKVOFQON 854
Qy 767 ENVSKLDVLEKLYEFLNLSYICHKYLIVSHSTMNEKILKOYITTEBESKLSQCDLIL 826
Db 855 TN--LRVIOKNOE--ETTSIS-----KITVNPSE-----ELF 883
Qy 827 FNIONNIVMYSMPDLSNLSLSQLEMEIYEKEMVNCYLKLDNDKIKNLEBAKRYSTSV 886
Db 884 SDNENNF--VFQVANNERNNALGNKTELHETDLTCVNEPIFNKSTMYLVXGDIQDAQY 941
Qy 887 KTLSSSSMOPLSLTPQDREVSANDTSHSTNLNLSLKFENILSLGKNKNIOE---L 942
Db 942 ---SIKRDLYVYVLAENKNKYQHIIKMTLQDLKSDISL--NIDKIPKNNNDYMKMAGL 996
Qy 943 IGOKSESENF-----YEKILKDSPTFYNESFTNPKSKA--DOINSLN-DES 985
Db 997 LGPISNHSFGSGFTPAKSKETIKLSHNTKSKMFKDIEBOYPTSLACVEIVNTIALDQ 1056
Qy 986 KKKLEEDINKLKTLOLSF--DLNKKYKLELRLFPDK-----KTVGKKYMOIKKLT 1036
Db 1057 KKLKQPSINIVYSAHLQSSVAVSDCKNSH-ITPQMLFSGKODNSNHNLTSPQAKETIELS 1115
Qy 1037 LKKEOLES--KLNSLNPKHVLQ--NFSVFENK-----KEAEI----- 1071
Db 1116 TILBESGQPFOTQPRKPSYILOKSTFEVPEQOMTILKTTSECHDADHIVMNPASIQ 1175
Qy 1072 AETENTLENTLILKHYGLVK-----YNGESSPLKLTSEESI 1110
Db 1176 VDSKQFEGTVEIKKFKAGLKNDCNKSASGLYLDENEVEGFGRFSAHGTKL-NVSTEL 1234
Qy 1111 QTEDNYASLENFYVLSLEGLKLDNLNEKKKLSYSSGLHLILAELEVIKKNKYTGNS 1170
Db 1235 Q-----KAVKLFSDE-NISEETSAVHPISLSSSKCHDSVSM---FIEEHNDKT 1282
Qy 1171 PSENN-----TDVNNALSEYKKFLPECTDVATVVSSEGSQDLE---QS 1210

Db 1283 VSEKNKCOLLIÖNNIEMTGTGFEVEITENYKR-NTENEDNKYTAASRSHNLEFGSDS 1341
Qy 1211 QPKKRASTHVGAESNTITTSQVNDDEVYIIVPIFGESEEDYDL-----G 1257
Db 1342 SKNDYVCJHKDETDLLFTDQHNHICLSQGFKKEGNTQIKEDSLDTFLEVARAOEACHG 1401
Qy 1258 QVVTGEAVTPSYIDNLSIKIENEYEVLYLKLAGVYRSRKLQENNVMPF-----N 1308
Db 1402 MTSNKEQÜLATYTEONIKDEF--SDTFPQTASGKNISYAKELFNKIVYFPQKPEELHN 1459
Qy 1309 VVVKDILNSRFKRENFKNVL--ESDLIPYKDLTSS-----NYVKDPYKFLNEKRD 1359
Db 1460 FSLNSELHS--DIRKNKMDLSYEETDIYKHILKESVYVGQNDLY-----TFQOQPERD 1513
Qy 1360 KFLSSYNTKDSIDPDINFANDVLYKILSEKY--KSDLSIKKYINDK-QGENE--- 1412
Db 1514 E-----KIKED-----TILGFHPASGKRVIAKESLDKVNLEPDEKQGTSEITS 1558
Qy 1413 -----KYLPLNIEETLTKTY-----NDK-----IDLFY----- 1436
Db 1559 FSHQNAKTLTKREACKDLELACETIEITPAAPCKEKQMSLNDKNLVSIETVYPPKLLSD 1618
Qy 1437 -----KYLPLNIEETLTKTY-----NDK-----IDLFY----- 1438
Db 1619 NLCROTENLTKTSKSIPLKVKVHENVKEKTAQSPATCYTNQSPYSIENSALAFYTSCKRK 1678
Qy 1439 -----LEAK-----VLATYTB-KSNVEYKIKELNYLTQDK 1469
Db 1679 TSVQSOTSLLEAKRWLREGIFDQPERINTADVGNLYENNSNSTAENDKHNLEBKQDT 1738
Qy 1470 -LADFKKNNFYGIADLSTDYNNHNLTKF-FLSTGM--VFENILA-----KTVLSNLD 1518
Db 1739 YLSNMSMSYSYHSD--EYVDSRLSKNLKIDSGIEPLKAVDEOKNTSPKSVISNVKD 1796
Qy 1519 GN-----LOGMLNISOH 1530
Db 1797 ANAYPQTVNEDICVEELVYTSSPCKNKNAAIKLSISNMFVGPAPRIASGKIRLCSH 1856
Qy 1531 QCVKK-----QCPONSGCPFHLDRECKCLLWYKQEGCK 1566
Db 1857 EYIKRVKIDDSFKVIFTKENNEKSKICQYKIMAGCYEALDSDSD--TLNLSLNDNDC 1913
Qy 1567 VENPPTC-----NENNGCDADAKCTEEDSGSNGKKTICECT-----KPSYP 1610
Db 1914 SMHSKIVPADIOSEELQHNQMSGLKYSKISPODVSLFTSDI-CKCSIGKRLHSYSSA 1972
Qy 1611 LFDGIFCSSS 1620
Db 1973 NTCGIFSTAS 1982

RESULT 39

US-08-603-753D-4

; Sequence 4, Application US/08603753D

; Patent No. 5891857

; GENERAL INFORMATION:

; APPLICANT: HOLT, JEFFREY T.

; APPLICANT: JENSEN, ROY A.

; APPLICANT: PAGE, DAVID L.

; APPLICANT: KING, MARY-CLAIRE

; APPLICANT: SZABO, CSILLA I.

; APPLICANT: JETTON, THOMAS L.

; APPLICANT: ROBINSON-BENION, CHERYL L.

; APPLICANT: THOMPSON, MARILYN E.

; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2

; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ARLES A. TAYLOR, JR.

; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER

; CITY: DURHAM

STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603/753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HS043746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION: regulatory effect on growth of human mammary cells.
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
TITLE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-08-603-753D-4
Query Match 3.4%; Score 288.5; DB 2; Length 3418;
Best Local Similarity 18.6%; Pred. No. 5.4e-07;
Matches 382; Conservative 287; Mismatches 718; Indels 663; Gaps 92;

QY 172 EEINELLKLNFEYDILLRAKLANDVCANDYQI---PENIKIRANELDLKIVFGYRKP 227
DB 327 KTRKKIIFHEAN-----ADECEKSKNQVKEKSYFSEVBPNDPLDSNV-AHQRP 375
QY 228 LDNIKDNVGMEDYIKKKNKTININELIESKTTIDKNNNAKKEEKKLYQAQVDLSI 287
DB 376 FEESGDSKISK-----EVPSSLACWSQSLTSLGNGAQMEKIPILLHSSCDQNT 423
QY 288 YNQLLEFANHLISVLEKRIIDTLKKNENIKELDKINEI--KNPPANGNPNTLDKNNK 346
DB 424 SEKDLDLTEN-----KRRKDFLTSEN---SLRISSLPKSEKPLNDETVYNN-RDEQH 473
QY 347 IEER-----EKEIKELAKTIKFNIDSLF---TDPLELEY 378
DB 474 LESHTECILLAVKQASIGTSVAVASSFOGIKKSIPIRIRHSPKPEFNASGCHMTDP-NFKKE 532
QY 379 LREKKNIDISAKVEIK-STEPNEYPNQVITPLSYNDINNALNELNSFDLDINPDTYK 457
DB 533 TEASESGLEIHTVCQKEDSLCPNLIDNG-SWPATYTONSVALKMAGLISLTL----- 583
QY 438 EPSKNITYTONERKKFNEIKIEKKIIEKKIESDKSYEDRSKSLNDITKEYEKLNEYD 497
DB 584 -----KKTINKFIYAIHDETFYKGGKTIKPKQKS-----ELIN--CS 617
QY 498 SKFNNNIDLTNFEKMMGKRYSYVEKLTJHNTFASYEN-----SKNLEKLTKA 546
DB 618 AOFEPANA-----FPAPL-----TFANADSGLLHSSVKRSCQNSQNSEPPTLS 658
QY 547 LKYMEDYSLR-----NIVYEKELKYKULISKIENIEITVENIKKDEQLEFK 596
DB 659 LTSSFGTILKRCGRNETQSNNTVAISODLYKEKCNK-----EKLOFLITPEADSSCLQ 713
QY 597 ITRDENKPDKEILEVSDIVK-----VOQKVLMLNKIDELKLTQLILKVELKHN 647
DB 714 EGQCENDPKSK--KVSIIKEVLAACHPYQHSKVEISDIDPOSQSL----- 760
QY 648 HVDNSYKQENKQEPYLIYVLAKEIDKLVMPVYESLINEKKNKITEGOSDNSEPTG 707
DB 761 -----YDHENAST---LILPTSKDYLSNLY-----MISGRKESYKMSDKLKNNEESDV 807
QY 708 EITGQATTKPQOQAGSLBEDSVQAQAOBOKAOPVPVVPVEKKOVPPTAPVNN-NKT 766
DB 808 ELTKNIPMEKNQ-----DVCALNENYKN---VELLPKYMVAVSPSRKVOEN 854
QY 767 ENYSKLDYLEKILEYFLTSYICKYILVSHSTNMEKILKOYKITEEESKLSSCDPLDL 826
DB 855 TN---LRVIOKNOE--ETTSIS-----KITVNPDSK-----ELF 883
QY 827 FNIONNIPVMYSFDSLNNLSQLFMEIYERKENCNLKYLKNDKIKNLEAKKYSTSV 886
DB 884 SDNENNF--VFQVANNERNNALGNTKELHETDLTCVNEPIFKNSTWLYGDTGDKQATOV 941
QY 887 KTLSSSSMQPLSLTPPOKPEVANSANDTSHSTNINNSLKLFEINLSLCKKNNIYOE---L 942
DB 942 ---SIKKDLYVLAEEKNSVKQHIKMTLQODLKSQDISL--NDKPIEKNDYNNKKNKAGL 996
QY 943 IGOKSSSENF-----YEKILKSDTFYNESEFTNPFYKSKA--DDINSLN-DES 985
DB 997 LGRISNHSFGSFRITASNKELIKSEHNKSKMFKFKEIQGYPLSLACVEIVMTALDNG 1056
QY 986 KKKKLEEDINKLKTLOLSF---DLYNKYKLKLERLFDK-----KTVKRYMQIKKL 1036
DB 1057 KKLKSPQISINTVAHLOSSVAVSDCKNSH--ITPOMLFSEKODFNSNHNMLTPSOAAEITYELS 1115
QY 1037 LKKEQLES--KLNLSNPKPVLQ--NFSYFNNK-----KEAEI----- 1071
DB 1116 TILLESQSOEFQFRRPSYIILQKSTFEVPENDQITIKTSECRDLDLHINNAPSIGQ 1175
QY 1072 AETENTLENTKILKLBKGGVKK-----YNNSESSPLKTLSESI 1110
DB 1176 VDSKQEPGEIVETKRRFAGLLKNDCKNSAGYILTDENVGFRGFIYSAHGIKL-NVSTFAL 1234

```

01 1411 QTEENVASLENKRVLSKLGKTLKDNLTNEKKKSLYSSGHLHLIAELKEVIRKKNTGNS 1170
02 1235 Q-----KAVKLSDDIE-NISEETSAEYHPISLSSSKCHDSVSK--FKTENHDKT 1282
03 1171 PSENN-----TDVNNALBSEYKFKLEPGTDVATVSESQSDFLE---QS 1210
04 1283 VSEKNNKCOLLIÖNNIEMTTGTFFVEIEITENYKR-NTNEDDNKVTYASRNSHNEFPQSDS 1341
05 1211 QPKRASHTVAGAESNTITTSQNVADDEVDDIYVPIRGESEEDYIDL-----G 1257
06 1342 SKNDTVCIHKDETDLETFDQHNICLKSQAFMEGNTQIKREDISDLTFLEVARAKAEGACH 1401
07 1258 QVVTGEAVYPSYIDNILSKIEENEYEVLYKPLAGYRSJKOLENNVMFP-----N 1308
08 1402 NTSNKEQLTATEQNIKQFET--SDTFPQTASGKNISVAKELFNKIYVNFDDKPELHN 1459
09 1309 VAVKQILNSRFKKRENFKNVL--ESDLPYKDLTSS-----NYVVKDYPKFLNKEKD 1359
10 1460 FSLNSELHS-DIRKKMDILSYEEDIDYHKHLKESVPVGTGTONOLV---TFQGGPER 1513
11 1360 KFLSSYNYTKDSIDPDINPANDVIGYKILSEKY---KSDLSIKKYINDK-OGENE-- 1412
12 1514 E-----KIKEP-----TILGPHTSAGKKVYKAKESLDYVKKMLPDKBEGTSEITS 1558
13 1413 -----KYLPLNNIETLYKYV-----NDK-----IDLFY----- 1436
14 1559 FSHQWAKTLKYREACKDELACETIEITAPRCKEMQNSLINDKNLVSIEYVVPKLLSD 1618
15 1437 -----IH----- 1438
16 1619 NLCTROTENLKTSKSIFLKYKVHNEVEKEFPAKSPATCYTNOFPSYIENSALAFYTSCSRK 1678
17 1439 -----LEAK-----VLNTYVE-KSNVEYKIKELANTLKTIOQK 1469
18 1679 TSVQSTSLLEAKMLREGIFDGOPEIRINTADYVGNLYEENNSSTJLAENDKNHLSKQDT 1738
19 1470 -LADRRKNNPFQGIADLSTDYDHNHNLTK-FLSTGM--VFENLA-----KTVLSHLD 1518
20 1739 YLSSNSMSNSYSYHSD--EYVNDGYLSKKNKLDGSLIEPVALKVNDQKNTSFKSVISVWK 1796
21 1519 GN----- 1530
22 1797 ANAYQTVWEDICVELVYSSSPCKKNNAIKLISNSNNFVEGPAFRIRASKIRLCSH 1856
23 1531 QCYKR-----QCPONSCFPHLDERECCCLLNTKQEGDKC 1566
24 1857 ERIKIKVDIFDTDSFKVIEKENNEKSKIOQTKIMACGYELADESD--ILHNSLMDDEC 1913
25 1567 VENPPTCC-----NENNGCGDADAKCEEESGSGSKGRIICECT-----KPDYSP 1610
26 1914 SMHSKRVAPDIOSEELIÖHNNÖMSGLEKYSKITSPCDVSLETSDI-CKCSIGKLKHSVSSA 1972
27 1611 LFDGIFCSSS 1620
28 1973 NTCGIFSTAS 1982
29
30 RESULT 40
31 US-09-099-753-4
32 ; Sequence 4, Application US/09099753
33 ; Patent No. 6149903
34 ; GENERAL INFORMATION:
35 ; APPLICANT: HOLT, JEFFREY T.
36 ; APPLICANT: TENSEN, ROY A.
37 ; APPLICANT: PAGE, DAVID L.
38 ; APPLICANT: KING, MARY-CLAIRE
39 ; APPLICANT: SZABO, CSILLA I.
40 ; APPLICANT: JETTON, THOMAS L.
41 ; APPLICANT: ROBINSON-BENION, CHERYL L.
42 ; APPLICANT: THOMPSON, MARIYLN E.
43 ; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
44 ; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
45 ; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.

```

```

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
STREET: BOULEVARD
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,753
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEX: (919) 419-0383
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: unknown
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
TITLE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-09-099-753-4
Query Match 3.4%; Score 288.5; DB 4; Length 3418;
Best Local Similarity 18.6%; Pred. No. 5.4e-07;
Matches 382; Conservative 287; Mismatches 718; Indels 663; Gaps 92.

```

```

OY 61 AVTTSPGSKGVSAGSGGVSAGSVAS--GGSVASGGSVAGG-----SGNSRRTNPS 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 VERPHDTAVKVSFFSNHDESLKKNDRFIASVDTSEMENTNOREASHGFGKTSGMSFVN-- 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 115 DNSSDSADKSYADLKHVRVRYLLTJIKELYPOLFDLTNH-MLTCDNIHGEKYL--IDGY 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 -SCKDHIGKSMRVN-----LEDEVEYETVADTSEEDSFSLCFSCKRTKMLQWRTS 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 172 EELNELLYKINFEDLLRAKLNDCANDYCOI-----PENKIRANELDVKKRYFGYGRK 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 KTRKKLJFHEAN-----ADECEKSKNQVEKYSFVSEVEPNDDPLDSNV-AHQKP 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 228 LDNIKNVGMEDYIKRNNKTJENINELIEESKRTIDKNKNATKEEEKKLYOAOYDLST 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 FEGSGDKISK-----EVVPSLCEMSQULTSLGNGAQMEKIPHLHSSCDQNI 423
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 288 YNKQLEAHNLISVLEKRIIDLTKNENIKELDKINEI-KNPPRANSNTPTNLLDKNKK 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 SEKDLIDTEN-----KRRKDLTSEN--SLPRISLPSKSEKPLNEETVVK-RDEEQH 473
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 347 IEERH-----EKEIKEIKTIKFNIDSLF---TDPLELEY 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 LBSHTOCLIAVKAISGTSFVASSFOGIKKSTIRRESPEETNASTSGHMTDP-NFKKE 532
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 379 LREKKNIDISAKVETRE-STEPNEYNGVTPRLSYNDINNALNELNSFGLINPEDYTK 437
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 533 TEASESGLEIHVYCSQKEDSLCPNLIDNG-SMPATTQNSVALKNGLISTL----- 583
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 438 EPSKNITYTNEKFKTNEIKETIKIKIKIESKKSVEYEDRSKSLNDITKEKYLNEYD 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 584 -----KKTNKFTYAIHDETFYGGKIPKOOKS-----ELIN-CS 617
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 498 SKFNNNIDLTNFEKMGKRSYKVEKLTJHNTPASEYEN-----SKHMLEKTKRA 546
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 618 AOEPAAN-----FEAPL-----TFANADSGLLHSVYKRSQONSEPTLS 658
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 547 LKYMEDYSLR-----NIVERELKYKYLISKIENEIEFTLVENIKKDEOLEKK 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 659 LNSSFETILKCSRNFTCSNNTVVISODLDYKREAKCNK-----EKLOLFIPEADSLSCLO 713
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 597 ITKDEKKPBEKILEVSDIYK-----VOVOVKVLMNKIDELKKTQOLLKKNVELKHN 647
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 714 EGGEENDPYSK-KVSDIKIEEVLAAACHPVOHSKVEYSDTDFOSOKSL----- 760
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 648 HVPNSYKOEKOBEPYLYLKEKIDKLKVPMPVESLINEEKNKITEGOSDSESTEG 707
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 761 -----YDHENAST--LILPTSKDVLNLY-----MISGKESYKMSDKLCKNNYESD 807
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 708 EITGQATTFRGOAGSALBEGDSVOAOAOEOKAOPVVPVPEAKAOVTPPAV-NKT 766
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 808 ELTKNIPMEKNQ-----DVCALNENYKN--VELLPPEKYMVASPSRKYOFONQ 854
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 767 ENVSKIDYLEKLEFINTSYICHKYLIVSHSTNNKILKQITKEESKSLSCDPLDLL 826
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 855 TN--LRVJOKNOE-ETTSIS-----KITVNDSE-----ELF 883
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 827 FNIOANNIPWYSMFDLSLNSLSOLFMEIYEKENVCNLYLKLDNDKIKNLLEAKKYSTV 886
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 884 SOENENF--VFOYANERNNALGENTKELHETDLTCVNEPIFKNSTVILYDITDOKATOY 941
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 887 KTLSSSSMOPLSLTPODKPEVNSANDTSHSTNUNNSIKLEFENILSLGKNKIYOE---L 942
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 942 ---SIRKDLVYLAEBKNKNSVKOHIMKMTLQODLKSDISL--NIDKIPEKNNDMYMNKAGL 996
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 943 IOKSSENF-----YEKILKODSFYNESFTNFYKSA--DDINSIN-DES 985
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 997 LGPISNHSFEGSFRJASNNKIKLEHNHKKSKMFEIDIEOYPTSLACVAVITVALDNO 1056
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 986 KKKLEEDINKLKTLOLSP---DIYNKYKLTLEPLDK-----KTVGKYMOIKT 1036
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1057 KKLKSKQSINTVSAHLQSSVYVSDCKNSH-ITPOMLFKQDFNSNNLTPSOAKAETIELS 1115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OY 1037 LKEOLES--KINSLNPKHVIO--NEFVENK-----KEAET----- 1071
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1116 TLEESGOFEEFTQFRKPSYIIOKSTFEVENOMTILKTSECBRADLIVMANSIQO 1175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1072 AETENLENTKILKHKYGLV-----YYGESSPLKTLSEESI 1110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1176 VDSSKQEGEVEITKRRFAGLKNDCNKSASGYLTJENENGFRGFYSAGTKL-NVSTEAL 1234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1111 QTDONVASENFVLSKLECKLNDNLNLEKRLKSLYSGLJHNLIAELKVKYKNYTGNS 1170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1235 Q-----KAVLFSDIE-NISEETSAEVHPISLSSKCHDSVGM--FKTEINNDKT 1282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1171 PSENN-----TDVNNALSEYKFLPEBGDTAVVSESGDYLE---OS 1210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1283 VSEKNNKCOLLIIONNIEMTGTFEVEETENYKR-NTENENKRYTASRNSHNLEPFGSDS 1341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1211 QPKKPASTHVGASNTITTSQNVNDEVDVYIYPIGSEBEDYDL-----G 1257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1342 SKNDVYCIHDEFTDLFTPDHNLCLKLSQGFMEKGTQIKEDLSDLFLEVAKAQCACHG 1401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1258 QVVTGEAVTPSVIDNLSKLENEYEVLYLKPLAGVRSKLOENNVMTF-----N 1308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1402 NTSNKEQLTATKTEQONIKOPEI--SDTFQOTASGKNISVAKELFNKIVNFQDKPRELHN 1459
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1309 VNVKDLINSEFNKRENFKNVL--ESDLIPYKOLTS-----NYVVKDYPKFLNKEKD 1359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1460 FSLNSELHS--DIRKKNMDLSYEEDIDYKHKILKESVPVGTNOQLV--TFQGGPEBD 1513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1360 KFLSSYNYIKDSIDTQINFANDVLYGYKILSEKY--KSDLSIRKYYNDK-OGENE-- 1412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1514 E-----KIREP-----TLGFRHTASGKVKYIAKESLDVYKKNLFDEKEGTSEITS 1558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1413 -----KYLFPNNIETLYKTV-----NDK-----IDLFY----- 1436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1559 FSHQWAKTLKRYRACDDELACETIETLAPCKCKEMQNSLNNDKNLYSIETVVPKLLSD 1618
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1437 -----IH----- 1438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1619 NLCROTENLTKSIFLKYVHVENHVEKETAKSPATCYTNOPSYIENSALAFYTSCSRK 1678
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1439 -----LEAK-----VLANTYE-KSNVEYKIKELNYLTKTIQDK 1469
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1679 TSVSOTSLKAKKWLREGIFDQOPERTNADYGNVILYENNSSTIAENDKNLSEKQDT 1738
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1470 -LADFKNNNFVGIADLSTDYNNHNLTK-ELSTGM--VEENLA-----KTVLSNLLD 1518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1739 YLSNMSMSNGYSYHD--EYVNDGYLSKKNKLDGIEPVLKVEDQKNTSFSKVISVYKD 1796
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1519 GN-----LOGMLNISQH 1530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1797 ANAYVPTVNEDICVEELVYSSPCKKNMAIKLISNSNNEFVGPAPRIASGKIRLCSH 1856
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1531 QCVKK-----OCPONSGCFRLDERRECKCLANYKQEGDKC 1566
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1857 ETIKKYKIDITDSFKYIKENNEKSKICOTKIMAGCYELADSEB--ILHNSLNDDEC 1913
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1567 VENPNPTC-----NENNGCDADAKCTEEDSGSNGKITECT-----KPDSTP 1610
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1914 SMHSKVFADIOSEELIOHONMSGLEKYSKISPCDVSIETSDI-CKCSIGLHKHSVSSA 1972
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1611 LFDGIFCSSS 1620
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1973 NTCGIFSTAS 1982
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 41

US-08-986-106-4

; Sequence 4, Application US/08986106

; Patent No. 6177410

; GENERAL INFORMATION:

; APPLICANT: HOLT, JEFFREY T.

; APPLICANT: JENSEN, ROY A.

; APPLICANT: KING, MARY-CLAIRE

APPLICANT: STEINER, MITCHELL S.
APPLICANT: ROBINSON-BENION, CHERYL L.
APPLICANT: THOMPSON, MARILYN E.
TITLE OF INVENTION: THERAPEUTIC METHODS FOR
TITLE OF INVENTION: PROSTATE CANCER
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986,106
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
FEATURE:
NAME/KEY: BRCA2 protein
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
TITLE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box domain at
RELEVANT RESIDUES IN SEQ ID NO: amino acids 3334-3344
US-08-986-106-4

Query Match 3.4%; Score 288.5; DB 4; Length 3418;
Best Local Similarity 18.6%; Pred. No. 5.4e-07;
Matches 382; Conservative 287; Mismatches 718; Indels 663; Gaps 92;

QY 7 LCSFLP-----FIINTQCTH--ESYQELVKKLEALEDAVLIGYSLFQEKKNVLTNGTSGT 60
DB 160 VCGSLFHTPKFVKGKQTPKHSSESAGAEVDPDMSSSSSLATPPPTLSVTLVYRNEASEET 219
QY 61 AVTTSPGSKGVSAGSGSVASGVSAS--GGSVASGVSAGS---SGNSRRTMPS 114
DB 220 VFPHTTANVSYSFNSHDESILKKNDRTIASYTDSENTNORERASHGFGKTSQNSKFN-- 277
QY 115 DNSSDSAKSYADILKRVRYNLTITKELKYPQLFDLTNH-MLTLDNTHGKYL--IDGY 171
DB 278 -SKDHICKSPNV-----LEDEYETVYDTSSEDSFSLCFSCCRKKNLQKVRTS 326
QY 172 EFINELLYKLNFPDILAKINDVCANDYCOI---PNNKIRANELLVYKLVGKYPKP 227
DB 327 KTRKKIFHEAN-----ADCEKSKNOVKKEKYSFSEVEPNDTDLPSNV-AHOKP 375

QY 228 IDNIKDNVKNMEDYIKKNNKTJENINELIEESKKTIDKNNAKYEEKKKLYOAOYDLSI 287
DB 376 FESSGDKISK-----EYVPSLACESSQLTLSGLNQAOKKPIPLHLISSCDONI 423
QY 288 YNKOLEAHNLISYLEKRIIDLTKKNNIKELDKINEI-KNPPANSNGTPTLLDKNK 346
DB 424 SEKDLDTEN-----KKKDDLTSEN---SLPRISLPSKSKPIINEETVYVK-RDEQH 473
QY 347 IEFH-----EKEKELAKTIKFINDLF---TDPLELEY 378
DB 474 LESHTDCLAAVKAQALISGTPVASSPQIGKSLFRIRRESPKETFNASFQHMIDP-NFKKE 532
QY 379 LREKKNIDISAKVETKE-STEPNEYPNGVYPLSYNDINNALNELNSFGDLINPDYTK 437
DB 533 TEASESGLEIHTVCQKEDSLCPNLIDG-SMPATTONSVALKNAAGLSTL----- 583
QY 438 EFSKNITYDNERRKPIINEIKKIKIEKKIESDKSYEDRSKSLNDITKEYEKLNEYD 497
DB 584 -----KKTKNKFIVAHDETFFYKGRKIPQOKS-----ELIN-CS 617
QY 498 SKFNNDLTNFKKMGKRYKVEKLTNHNFAEYEN-----SKHNLEKTKA 546
DB 618 AOEANA---FEAPL-----TFANADGGLHSSVKSCSQNDSEPTLS 658
QY 547 LKYMEDYSLR-----NIVVEKELKYKNLSIKIENEIETLVENIKKDEOLFEEK 596
DB 659 LTSSFGLTKRCSRNQTSNNNTVISQDDLYKAKCNK-----EKQLFTTPBADSLSCQ 713
QY 597 ITKDKNKPKDEKILEVSDIVK-----VOYOKVLLNNKIDELKKTQILIKVNLKNI 647
DB 714 EQCENDPKSK--KVSIDIKIEVILAAACHPVQSHKVEYSDTQFOSOKSLD----- 760
QY 648 HVPNSYKQENKQEPYLLVYKKEIDKLKVPMPKVESLINBEKKNIKTGGOSNPSPTG 707
DB 761 -----YDHNAST---LITPTSKDVLNLY-----MISRGESYKMSDKLKNYVESDV 807
QY 708 EITGQATTKPGQAGSALEGDSVQAQAOBOKQAPVVPVPEAKAQPTPAPVY-NKT 766
DB 808 ELTKNIIPMEKNQ-----DYCALNENYKN---VELLPREKMYRASPCKYQAFON 854
QY 767 ENVSKLDYLEKLYEFLNTSYCHKYLIVSHSTANEKILKQYKTYEESKSSCOPLDL 826
DB 855 TN---LRVIOKQOE---ETTSIS-----KITVPDPSE-----ELF 883
QY 827 FNIGNNIIVMYSMPDLSNLSQLFMETYEKEMVONLKKIKNDKIKNLEBAKKVYSV 886
DB 884 SDNENN--VFQVANERNNLALGNTEKLEHETDLCTVNEPIFNKSTWVLXGDTGDQATOV 941
QY 887 KTLSSSMQPLSLTPQDRPEVSANDTSHSTNLNLSLKLFENILSLGKNKNYQOE---L 942
DB 942 ---SIKKDLVYVLAENKNNSYKQHIKMTLGQDLKEDISL--NIDKIPPEKNNDYMKKAGL 996
QY 943 IGOKSSENF-----YEKILKSDTFYNESFTNFKSKA--DDINSIN-DES 985
DB 997 LGPISNHSFGSGSFRASNKEIKILSEHNIKKSMFKEIDIEQYPTSLACVEIYNTIALNQ 1056
QY 986 KRKKLEEDINKLKTLOLSF---DLXNKYKLELELPKK-----KVVGKYMOKIKLT 1036
DB 1057 KKLSPQISINIVYSAHLQSSVYVSDCKNSH-ITPQMLFSKODPNSNHNLTPOKAEITELS 1115
QY 1037 LLKEOLES--KLNSLNKNKHYLQ--NFSVFFPKK-----KEAEI----- 1071
DB 1116 TILESSEQFETQPRKRSYILQKSTFVPEPNQMTILKTTSECDADLHYIMNAPSIG 1175
QY 1072 AETENTLENTKILKHYGLVK-----YNGESSPLKTLSEST 1110
DB 1176 VDSKQFEGTVEIKRFPAGLTKNDCKNSGASYLTEDENEGFRGFYSAGTKL-NVSTBAL 1234
QY 1111 QTEDNYASLBNPKVSLKLEGLKQNLNPKKKLSTLSGLHLLAEKLVKKNKNTYONS 1170
DB 1235 Q-----KAVKLFSDIE-NISETSAEVPHPISLSSKCHDSVSM--FKIENHNDKT 1282
QY 1171 PSENN-----TDVNNALESYKKFLPECTDAVAVVSESGSDTLE---QS 1210


```

Db 1283 VSEKNNKCOLILONNIEMTGTGFEVEETENYKR-NTENEDNKYTAASRSHNLEFDGSDS 1341
Qy 1211 QPKKPASTHVAGASNTTTSQNVDDVDYIYPIGESEEDDL-----G 1257
Db 1342 SKNDIVCINHEDDELFTDQHNCLKSGQFMKEGNTQIKLEDSLDFLEVARAQACHG 1401
Qy 1258 QVVTGAVPSPVIDNLSKLENEYVLYLPLAGVYRSKLENNMWT-----N 1308
Db 1402 NTSNKKQLTATKTEQNIKDEET--SPTEFOTASGKNISVAKELFNKIVNFDQKPEELH 1459
Qy 1309 VNVKDLNFKRNFKNVNL---ESDLIPKDLTSS-----NYVKKDPYKFLNKKRD 1359
Db 1460 FSLNSELHS--DIRKNKMDLSEETDIYVKKILKESVPGTGNOLV---TEGOGPERD 1513
Qy 1360 KFLSSTNYIKSIDTDINFIANDVIGYKILSEKY---KSDLSIKKTIINDK-OGENE--- 1412
Db 1514 E-----KIKEP-----TILGPHIASGKVKAKESLDKVKNLDEKEGTSITS 1358
Qy 1413 -----KYLPLNNIETLYKTV-----NDK-----IDLTV----- 1436
Db 1559 FSHQWAKTLKREACKDLDELACETIETAPKCKEMQNSLNDKNLVSITVVPKLLSD 1618
Qy 1437 -----IH----- 1438
Db 1619 NLCROTENKTSKIFLKVKYHENVETAKSPATCYTNSPYSVIENSALAFYTSCSRK 1678
Qy 1439 -----LEAK-----VLNLYTE-KSNVEYKIKELNYLKTIOK 1469
Db 1679 TSVSOTSLEAKKMLREGIFDGOPEINTADYGNITYENKNSSTIAENKNSHSEKQDT 1738
Qy 1470 -LADFKNNNFVGIADLSDYNNHNLTK-FLSTGM--VEENLA-----KTVLSNLD 1518
Db 1739 YLSNSSMSNSYSYHSD--EYVNDGYSLSKKNLDSGLEPVLKVNEDQKTSFSKYISVWK 1796
Qy 1519 GN----- 1530
Db 1797 ANNYPTVNEDICEVEELVYSSPCKNKNAIKISISNSNFEVGPAPRIASGKIRICSH 1856
Qy 1531 OCYK-----OCPONSGCFRHLDERECCCLLNKQEGDK 1566
Db 1857 ETIKKVKDIFTDFSKYIKENNENKSKICQTKIMAGCYEALDSDSD--ILHNSLDDEC 1913
Qy 1567 VEPNPTC-----NENNGCDADAKCTEEDSGSNGKITCECT-----KPDSTP 1610
Db 1914 SMHSVYFADIQSEELIQHNMNSGLEKYSKISPCDVSLETSDI-CKCSIGKLHKVSSA 1972
Qy 1611 LFDGIFCSS 1620
Db 1973 NTCGIFSTAS 1982

```

RESULT 42

```

US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MUL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label=X
; OTHER INFORMATION: /note="X" = M and N, or N"
US-08-290-919-4
Query Match 3.3%; Score 282; DB 1; Length 53;
Best Local Similarity 94.2%; Pred. No. 7.1e-09;
Matches 49; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1570 PNPTCENNGCDADAKCTEEDSGSNGKITCECTKPDSTPLFDGIFCSSN 1621
Db 2 PNPTCENNGCDADAKCTEEDSGSNGKITCECTKPDSTPLFDGIFCSSN 53

```

RESULT 43

```

US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992

```


OY	1273	ILSINENYEVL- ----KPLAGYRSJKOLENNV- --FENNVADILINSRENKEN	1324
Db	1101	LPMFLTSDYETSILVYDSRKGIRGCTKEALIEHTSHELVDAAFNVTMLITRSILITREF	1160
OY	1325	EKNVL-ESDLPYKDLTSSNVV- ----KOPYKELKEKROKFLSSY- --NYIKSDTIDT	1375
Db	1161	FYALITRYNYLPEEGSLDYDDYDNIWIEKSPNPIRCRVNIMRFLFYQWTRMYNEGPIJLI	1220
OY	1376	INFANDVIAGYKILISEKKYKSDLDISIKRYINDKO-GENEKYLPFLNNIEFLKTYVDKIDL	1434
Db	1221	LNPFK- ----MVSSEKIPAEEDLLK- INEKLINENK- ----EPVDPKQODSUSA	1285
OY	1435	FV- ----IHLEA- ----KVLN- --ITYEKSNEVEKIKELN- YL	1463
Db	1266	VWQTTKRDNKSPHIMSSSLPSSASSAFRLKLLKLLDDIPTYA- --TOLTIVEHLDYL	1322
OY	1464	KTIODKILAD- --FKNNNFVGIAIDLTVDYVHNHLLTRKFLSTGMVFNELKATVLSMLDGN	1520
Db	1323	RITMECELDRAWGTRKYCNMGSGSPNITRFLNANNTLTUNFVSHITVQADVKT- RSKL- ----	1377
OY	1521	LOGMLNISOH	1530
Db	1378	TOYEVTVAOH	1387

Db 1300 ETIELTAPKCKEMQNSLNNDKNVSIETVVPKLLSDMLCQTEHLKTSKSTFLK----- 1355
Qy 1091 LKYYNGESSPLKLTSEESIQTEDNYSLEN-----FKVLSKLEGLKLDNLEKKLSYL 1146
Db 1356 -VKVHENVEKETAKSPATCYTQSPYSYIENSALAFYTCSSRKTYSQSLLEAKK--WL 1412
Qy 1147 SSGHLHLIAELKEVKIKNNKYTQNSPSENTDVNNMLESKFKFLPECTDAVATYSESGSPT 1206
Db 1413 RRG-----IEDQPERINTADYGVNLYENNSN-----STIAEDKNH 1450
Qy 1207 LEOSQPKKPASTHVAESNTITTSQNVDEVDVILPIFGSEEDYDGLQVVTGEAVT 1266
Db 1451 LSEKQ-----DTYLSNMSNSYSYH-----SDEYVNSG----- 1480
Qy 1267 PSVIDNLLSKINEYEVLKPLAGVYRSLKOLENNMTFVNANKDIINSRKNRENK 1326
Db 1481 -----YLS-----KKNLDSGIEPVKKNVEDQKNTSFSK----- 1508
Qy 1327 NVLESOLIPYKDLTSSNVYKDPYKFLNKEKRDKFLSSVNYIKDSITDINFANDVLYY 1386
Db 1509 -----VISN--VKDA-----NAYQVYNEDICVEELVTS-- 1535
Qy 1387 KILSEKYSDDLSDIKKYYINDKOGENEKYLPLNNIET---LYKTVNDKLDIFYIHLEAKV 1443
Db 1536 ---SSPCKKNKNAIKLSTISNS-----NNEFVGPAPRIASGKI--VCVSHETIKK 1580
Qy 1444 LNTYTEKSANVEYKIKELATLTQDKLADFKKNNNFVGIADISTDYNNHNLTKELST-- 1501
Db 1581 VADITDTSFSKY--IKENNENKS--KICQTKIMAGCYEALDSEDLIHLNSLNDDECSH 1636
Qy 1502 GWFFENLACTVSNLDGMLNLSOHQCYKQCPQNSGCFRHLDEREECKCILLNYQ 1561
Db 1637 HNVFADIQSEEL-----LQHNQNSGLEKYSKISP-----CDVSLTSDICKCSIG--- 1682
Qy 1562 EGDKCVENPNPTC 1574
Db 1683 KLHKSVSAN--TC 1694

RESULT 46
US-08-639-501-2
Sequence 2, Application US/08639501
Patent No. 5837492
GENERAL INFORMATION:
APPLICANT: Tavligian, Sean V.
APPLICANT: Kamb, Alexander
APPLICANT: Simard, Jacques
APPLICANT: Couch, Jacques
APPLICANT: Rommens, Johanna
APPLICANT: Weber, Barbara
TITLE OF INVENTION: Chromosome 13-linked Breast Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Venable, Baetjer, Howard & Civiletti
STREET: 1201 New York Avenue N.W., Suite 1001
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 22204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,501
FILING DATE: 29-APR-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/585,391
FILING DATE: 11-JAN-1996

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/576,559
FILING DATE: 21-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/575,359
FILING DATE: 20-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/573,779
FILING DATE: 18-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Imnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-116802-04
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-639-501-2

Query Match 3.3%; Score 279.5; DB 2; Length 3418;
Best Local Similarity 21.0%; Pred No. 1.6e-06;
Matches 330; Conservative 225; Mismatches 569; Indels 449; Gaps 81;

Qy 187 LRAKALNDVCAN-----DYCQIPENLK---IRANBLDVLKLVFGY-----RKPL 228
Db 667 LKRCSENFTCSNNTVISQDLRYEAKCNKEKQLFITPEADSLSCLOEQCENDPKSKYV 726
Qy 229 DNKDNV-----GAME--DYIKNNKTI---ENINELI--EESKKT-----ID 264
Db 727 SDIKEEVILAAACHPVQSHSVESYSDTDFQSQSLDYDHENASTLLIPTSKDVLNLMWTS 786
Qy 265 KKNNAKREEEKKKLYQAYDLSYNNKOLEAHNLISVLEKRIPTYLTKNNENIK--ELL--D 320
Db 787 RGESEYKMSDKLK-----GNNYESVDLTKNI--PMKKNQDVCLANENKYNVELLPRE 837
Qy 321 KINEIKNPPANSNGTPTNLLDNKKIEHEKEIKEIAK--TIKENIDSLFTDPLELEYVL 379
Db 838 KYRWVSPSRKVOFN-QNTNL--RVIQKNOEETTSISKITVAPDSEELFSD----- 885
Qy 380 REKNKIDISAVETKESEPEPNPYGYPTSYNDINNALNE---LNS---FGDLINP 432
Db 886 NENNEVFOY-----ANERNNALGNTKELHETDL--TCVNEPIPKNSTWVLYGDTGK 936
Qy 433 FPTYKEPSKN-IYTDNERKKFINKIEKIKIE--KKKIESDKRSYEDR--SKSLNDITKEYE 489
Db 937 QATQVSIKKDLYVLAENK--NSVKQHITKMTLGQDLKDISLNDIKIPEKNNNDYNNKWA 994
Qy 490 KLLNELYDSKPNNNIDLTNFEKMMGRYKYKEKLTHTHTFVSYENSKNNLEKTRALKY 549
Db 995 GILGPISNHSFGGSF-----RTASNKTKLSEHNIRKSKMFKFQIEEDYPTSLAC 1044
Qy 550 MEDYSARNIVVERKELKYK--NLISKIENETLEY-----ENIKKEDEOLFPEK--IKDEN 602
Db 1045 VE--IVNTLALDQKLSKPSQINTVSAHLQSSVAVSDCKNSHTPOMLFSKODFNSNN 1102
Qy 603 -RPDEK--ILEVSDIVKVOVKVLLMKNKIDELKQTOLLKVELKHNHIVPNSY----- 653
Db 1103 LTPSQKAEITELSTLEESGSQF-----EFTQRRKPSYILQ-----KSTFEVPEKQMTILKT 1154
Qy 654 -KOENKQEPYVILVYK--ETIDKLVPKPYE-----SLNEEKKNIKTGQSDNSNP 703
Db 1155 TSEECRDADLHIVMNPISIGOVSSKQFEGTVEIKRKFAGLLKNCNKASAGVLDENEV 1214
Qy 704 STREGITGQAT-----TKPGQQA--GSALDEDSVQAQOEDQKQAPVPVPVPEAKAY 755
Db 1215 GFRGFTSAGHTKLANSTELAQKAVKLFSDIENISETSAEVH-----PLSSSSKHDSDV 1269
Qy 756 PTPPAPVNNKTENVSKLDYLEKLYEFLNFSYCHKYYILVSHSTMNKELIKQYK--ITKEBE 814

```

Db 1270 VSMFKIENHNDKTVSEKNNKCOL-----LIQNNIEMTGTGFEVEITENYKNNTNED 1321
QY 815 SKLSSC-----DPLDLEFNIONIPVMS-MF-----DSLH 844
Db 1322 NKTTAASRNHNLEFDGSSSKNDYVCIHKDEDDLFTDQHNICKLKSQGMKEGNQIK 1381
QY 845 NSISQL-FMEIYERKVCNLYKLKDNKIKNLEAKRVSTVKTSSSMOPLSLPOD 903
Db 1382 EDLSDLTFLEVARAKQACH-----GNTSNKEQL-TATKTQONIKDFETS----- 1424
QY 904 KPEVSANDOTSHSTNLSNKLFEENILSLKNNNIYELLGOKSSEMFYKILKDSPTF- 962
Db 1425 -----DTEFQY-----ASGNISVAKASEPNK--TVNFPQKPEELHNF 1461
QY 963 -YNEFTNFKSKADDIN-SLNDESKRRKLEEDI-----NKLKTLQ 1002
Db 1462 LNSELSDIRKNNMDLSTYETDVKHKILKESVPVGTGMQVTFQOQPERDEKIKPTL 1521
QY 1003 LSFELYNKKYKLT-----ERLFDR-----KTVGRYKMOIKKTLK 1039
Db 1522 LGFHTASGRKVKLAKESLDKRVKNLFDEKQGTSEITFSHQMAKTL-KYREACKDLELAC 1580
QY 1040 EOLE-----SKLSLNPKHYLQNFVYFENKKEAETA-ETENLENTKILKHYG 1090
Db 1581 ETEITAPACKEMQNSLNDKMLVSIETVPPKLSLNDLCROTENLTKTSIFLK----- 1636
QY 1091 LVKYNCESSPLTLSEESIQTEDNVASLEN-----FKVLSKLEGLKDNINLEKKLSYL 1146
Db 1637 -VAVHEVVEKETAQSPATCTCTNOSPYSVIENSALAFYTSOSRKTSVSOTSLERKK--WL 1693
QY 1147 SSGLHLIAELKEVYKNNYGTGSPSENNTDVNNALESYKFLPEGTVAIVVSESSDT 1206
Db 1694 REG---IFDQPERINADYVGVLYENNEN-----STIAENDKNH 1731
QY 1207 LEOSQPKRPASTVGAESNTITTSQNVDEVDVYIPIGSEBEDYDLQGVYTGAVT 1266
Db 1732 LSKQ-----DYLSNSSMSNSTYH-----SDEVNDSG----- 1761
QY 1267 PSYIDNLSKIENEYEVLYLKLPLAGYVRSUKOLENNVMTFNVKDIILSRFNKRENFK 1326
Db 1762 -----YLS-----KNLDSGIEVLKVNEDQKNTSPSK----- 1789
QY 1327 NVLESLLPYKDLTSSNYVVKDPYKFLNKKRDKFELSSVYIKDSIDTDFINPANDVGLGY 1386
Db 1790 -----VISN--VKDA-----NAYPQTVNEDICVEELVTS-- 1816
QY 1387 KIISEKYKSDLSIKKYINKOGENEKYLPLNNIET---LYKTVNKKIDLPVTHLEAVY 1443
Db 1817 ---SSPCKNNAIKLISLSS-----NNFEVGPAPFRASGKI--VCVSHETIKK 1861
QY 1444 LNTYEKSNVEVRIKELINYKTLTQDKLADFKKNNNFVGIADLSTDYVHNNMLTKFLST-- 1501
Db 1862 VKRIFIDSPFKV--IKENENKMS---KICQFKIMAGCEALDDSDILANSLDNDCEGTHS 1917
QY 1502 GWFENLAKTVLSNLDGNLQGMNLISQHCQVKKQCPQNSGCFHLDRECECKLNYKQ 1561
Db 1918 HKFFAIOSEEI-----LOHNNMSGLEKVSISP-----CDVSLSTSDICKCSIG--- 1963
QY 1562 EGDGCVENPPTC 1574
Db 1964 KLHKSVSSAN-TC 1975

```

RESULT 47
US-09-044-946-2
Sequence 2, Application US/09044946
Patent No. 6033857

GENERAL INFORMATION:
APPLICANT: Tavligian, Sean V.
APPLICANT: Kamb, Alexander
APPLICANT: Simard, Jacques
APPLICANT: Couch, Fergus

```

APPLICANT: Rommens, Johanna
APPLICANT: Weber, Barbara
TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESS: Venable, Baetjer, Howard & Civiletti
STREET: 1201 New York Avenue N.W., Suite 1001
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 22204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/044,946
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/639,501
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/576,559
FILING DATE: 21-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/575,359
FILING DATE: 20-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/573,779
FILING DATE: 18-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-116802-04
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-044-946-2

```

Query Match 3.38; Score 279.5; DB 3; Length 3418;
Best Local Similarity 21.0%; Pred. No. 1,6e-06;
Matches 330; Conservative 225; Mismatches 569; Indels 449; Gaps 81;

```

QY 187 LIRAKLNDVCAN-----DYCOIPFNLK-----IRANELDVAKLIVEGY-----RKPL 228
Db 667 LRKSRNMTCSNNTVISQDLDYKFAKCNKEKQLFTTPPADSLSCLOEGCENDPKSKVY 726
QY 229 DNIKDNV-----GKME--DYIKNNKTTI---ENINELI--DESKT-----ID 264
Db 727 SDIKKEVLAACHPVQSKSEYSDTPQSQSLKYLDHENSILILPTPSKDVLSNLVMS 786
QY 265 KKNKATKEEEKKLYAQYDLSTIYNKQLEBAHNLISLEKRIDTLKKNEMK--ELL--D 320
Db 787 RGKESYKMSDKL-----GNYESDVELTKNT--PMERNODVCALNENYKNNVLELPE 837
QY 321 KINIKNPPRANGNPNTLLDKNKKILEHKEKEIKIAK-TIKNIDSLTDPLELEYL 379
Db 838 KYMRVASPSRRVOFN-ONTNL--RVLOKQOETTSIKITVDPDSEELSD----- 885
QY 380 REKNKNIDISAKVETKESTPNEXVNGVTPVPLSYNDINNALNE--LNS-----FGDLINP 432
Db 886 NENNFVQV-----ANERNNLALGNTKELHETDL-TCVNEIPIFNSTMYLVGDTGDK 936
QY 433 FDTYKPSKN-IYDNERKKFINKIKIE-KKIESDKSYEDR-SKSLNDITKEYE 489

```

Db 937 QATQYSIKKDLVYLAEEK--NSVKOHKMTLGGODLKSDISLNDIKDIPEKKNNDYMNKA 994
Qy 490 KILNEIYOSKFNNDIDLNEFKAMGRYSYKVEKLTHTNTFASYSKNHLEKTKALKY 549
Db 995 GILGIGISHNSFEGSGF-----RTASNKEIKLEHNHKKSKMEFFKIEEOYPLSLC 1044
Qy 550 MEDYSLRNIVVEKELKYK--NLISKIENIEFLV-----ENKKEDEOFEEK--ITKDN 602
Db 1045 VE--IVNTLADNOKKLSKPOSINTVSAHLOSSVYVSDCKNSHITPOMLFSKODNSNHN 1102
Qy 603 -KPKDK--ILEVSDIVKVOVOKVLLMNKIDELKTQLILKNVELKHNHVPNSY----- 653
Db 1103 LTPSOAKETELTELLESGSGF-----EFTQFKPSYILQ-----KSTFEPENOMTILKT 1154
Qy 654 -KQENKQEPYILIVLK---EIDKIAKVPMPKE-----SLNEKKNKITGOSDNSP 703
Db 1155 TSEGRDADLHVIMNAPSIGOVSSKQEGVEIKRFRAGLKNCKNSASGYLTDENEY 1214
Qy 704 STEGEITGOAT-----TRPGOA--GSALEGDSVOAOEQKQOAPVPVPVPEAKAQV 755
Db 1215 GPRGYSAHGTYLANSTELQKAVLFPDIEINISETSAEVH-----PISLSSKCHDSV 1269
Qy 756 PTPAPVNNKTENVSKLDYLEKLYEFLMTSYICHKTYLLVSHSTNEKILKQYK--ITKEE 814
Db 1270 VSMFKIENHNDKTVSEKNNKCOL-----ILONNIEMTGTQFVEIEITENYKRNTEMED 1321
Qy 815 SKLSSC-----DPLDLFNIONNIPWYS--MF-----DSLN 844
Db 1322 NKTYTASRNSHNLFEFGSDSSKNDFVCJHKDETDLFTDQHNICLKLSQOFKKEGNTQIK 1381
Qy 845 NLSJOL--FMEIYEKEMWCNLYKTKDNDKIKNLEAKKYSTVKTLSSSQPLSTPGD 903
Db 1382 EBLSDLTLEVAKABEACH-----GTSNKEQL--TATKEQNIKDETS----- 1424
Qy 904 KPEVSANDTSHSTNLNNSLKLENILSGKNKNIYQELIGKSSSENEYKILKOSDTF- 962
Db 1425 -----DTEFOT-----ASGKNISVAKESFNK--IYNEFDQKPEELHNS 1461
Qy 963 -YNESFTNPFVKADDIN--SLNDESKRKLEEDI-----NKLKTLQ 1002
Db 1462 LNSELSHSDIRKKMKDILSTEEDIVYKHKLKLESVPVGTGNOLVTTQOGPERDEKIEPPL 1521
Qy 1003 LSFIDLNYKRYKTL-----ERLFDK-----KTVGKYKMOIKKLTLLK 1039
Db 1522 LGFHTASGKVKYIAKESLDKVNLFDEKQGTSEITFSHQMAKTL--KREACKODELAC 1580
Qy 1040 EOLE-----SKLNSLNPKHVLONTSVFPNKKKEAIA--ETENTLENTILKHKYK 1090
Db 1581 ETIEITAAPCKEKMONSINNDKNLVSIEFVVPKLLSDMLCROTENLKTSKSIFLK---- 1636
Qy 1091 LKYYNGESSPLKLTSEESIQEDNVASLEN-----FKVSKLEGLKLDLNLKKKLSYL 1146
Db 1637 -VKVHENVEKETAKSPATCYTNQSPSYENSALAFYTSCKRSTYSQTSLEAKK--WL 1693
Qy 1147 SSGHLHLAELKLVKKNKNTGNSPSENTDVNNALASYKFKLPBGTDVATVYSESGSDT 1206
Db 1694 REG---IFDQPERINTADYVGNLYENNSN-----STIAEDKHN 1731
Qy 1207 LEQSQPKRPASTHVAESNTITTSQNVDEVDVVIIPJFGESEEDYDGLGVVGEAVT 1266
Db 1732 LSEKQ-----DYLSNSSMSNSYSYH-----SDEYVNDG----- 1761
Qy 1267 PSVIDNIISKIENEYEVLYLKLAVYKSLKOLENNVTFVNVKDIILNSPKNRENK 1326
Db 1762 -----YLS-----KNKLGSIEPVLYKNVEDQKNTSFSK----- 1789
Qy 1327 NVLESDLIPYKDLTSSNVVVKDPYKFLNKKERKDKFLSSYNIKDSIDNDINANDVLYY 1386
Db 1790 -----VTSN--VKDA-----NAVQTVNEICVCEELVTS-- 1816
Qy 1387 KILSEKYSDDLSIKRYNDKOGENEKYLPLNNIET---LYKTVNDKIDLFVILHEARV 1443

Db 1817 ---SSPCKNKAIAKLSISNS-----NNEFVGPAPRIASGI-LVCVSHETIK 1861
Qy 1444 LNTTEKSNVEVKIKELNLTITQDKLADPKRNNFVGIADISTDYNNHNLTKPLST-- 1501
Db 1862 VQDIFDPSFKV--IKENNNKS---KICQTKIMAGCYEALDSEIDLHNSLNDCESTHS 1917
Qy 1502 GWFFENLAKTVSNLLDGMNLNSIQHCYKQCPQNSGCFRHLDEFECKCLLNTYKQ 1561
Db 1918 HXVFADIQSEEL-----LQHNQNSGLEKYSKISP-----CDVSLFTSDICKCSIG-- 1963
Qy 1562 EGDKCVENPNPTC 1574
Db 1964 KLHKSYSAN--TC 1975

RESULT 48

US-09-044-908-2
; Sequence 2, Application US/09044908
; Patent No. 6124104
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/044,908
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/639,501
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359
; FILING DATE: 20-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/573,779
; FILING DATE: 18-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Inuen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-116802-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO.: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-044-908-2
Query Match 3.3%; Score 279.5; DB 3; Length 3418;

;; FILING DATE: 16-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/329,154
;; FILING DATE: 25-OCT-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/161,907
;; FILING DATE: 02-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/985,321
;; FILING DATE: 04-DEC-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/429,791
;; FILING DATE: 31-OCT-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: INGOLIA, DIANE E.
;; REGISTRATION NUMBER: 40,027
;; REFERENCE/DOCKET NUMBER: OPHD-01763
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 705-8410
;; TELEFAX: (415) 397-8338
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2366 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-480-604A-10

Query Match 3.3%, Score 278.5; DB 1; Length 2366;
Best Local Similarity 18.5%, Pred. No. 1.2e-06;
Matches 347; Conservative 292; Mismatches 595; Indels 643; Gaps 92;

QY 121 DAKSADL-----KRVRYLL--TIKELKYPQLPDLTNHMLTCD--NIHGKYL----- 167
DB 51 DINSITDIYDTYKSGKNKALKKEKYLVELELKNNTLPVKNLH--FWIGQIND 109
QY 168 ----IDYEELINELLYKLNFEYD-----LLRAKLVDCANDYCOIPEFNKIRANELD 215
DB 110 TAINIYNQMDVNS--DYVNVWFYDINAFILNTLTKTVESAINDLLE--SFRENLDPRD 167
QY 216 VLKLIYVEYKRPDLNKKNGKMEYTKKN-----KTENINE 254
DB 168 YNK-----FFRRMELIYKOKNFYNYKAORENELIIDIIVKTYLSNEYKEIDELNT 223
QY 255 LIEESKTIIDNK--NATKEEKKKLYOAOYDLSYKOLEAHNLISYLE--KRITIKKN 312
DB 224 YIEESLAKITONSGNDVNFEEFKNGES---FNLYEDELVERMNLAAASDILIRISALKEI 280
QY 313 ENI---KELL-----DKINEIKNPFPANSANTPTLIDKNKKIEEH-----EK 352
DB 281 GGMVLDVMDLPGIOPDLFESTIEKPSVTVDFWEMTKLEAIKYYKEIYEYTSHEHDMDE 340
QY 333 EIKE-----IAKTIFENISLFTD---PLELEYLRKKKNKIDISAKVETKESTEPNE 402
DB 341 EVOSFFESVLASKDSKSEIFSLGDMEASPLEVK--IAFNSKGIINOGLISVYKDSYCNL 398
QY 403 YPNVG--TYPLSYNDINNALNELNSFGDLINPEDTKEPSKNIYDNRKKINIKI-- 458
DB 399 IVKQIENRKILNLSNLPAISDENDFTTTFIDISIAEAN--ADNR--FMMLGLGYL 454
QY 459 -----KIRIEKKIESDKSKSYEDRSKSLNDITKEYEKLEINLEYDSKFNINIDLTFN- 509
DB 455 RVGFEPDVKTITNLSCPEVAAVQDLDL---MFKESGMNHLEADLRNEISIKTINS 509
QY 510 ---EKAMGKRYSYKVEKLTHTNTFASYSKSNLEKTLKALKYME-----DYSLRN 557
DB 510 QSTDEMASLMSFDDARAK-----AOFEYKRN-----YFEGSLGEDDNDLDFS--QN 554
QY 558 IVEKELIYYNLISKINIEIT--LVENIKKD-----EEOFEKKI 597
DB 555 IYVQKE--YLEKISSLSRSEBGTIHYIVQOGDKIYEACNLFAPKTPYDSVLFQKNI 612
QY 598 TKDE-----NRPEKIEE-----VSDIVKVOVKR-----LLM 625

DB 613 EDSEIAYYNNPDGGEIQEIDKXKIPSIISDRKIKLTFIGHGKDEFENDIFAGFDVDSL 672
QY 626 NKID---ELKTIQLIKVNLKHNHVPNSYQOE--NKQEPYLIYLUKATIDLKAPMPKV 681
DB 673 TEIEAIDLAKEDISPKSIEI--NLGCMFPYSINVEETYPGKLKDKDISELMPSI 730
QY 682 --ESLI---NEEKKNITKGOSDNSPEPTEGETTGQATKPPQAGSALGDSVOAQOE 736
DB 731 SODSIIVSANQYEVARINSGRRELLDHSGE-- 761
QY 737 OKQADPPVPVPEAKAQPFPAPYNNKTNENYSKLDYLEKLYEPLNTSYICHKYLIVSH 796
DB 762 -----INKEESITK--DISKYEISPN----- 782
QY 797 STMNEKILKQYITKEESKLSLSCDPLDLFNIONNIPWYMFPSLNSLSQLEMEIYE 856
DB 783 -----KENKITVSK-----NLPGLSTLLOETIRNNSNSDIETEE 817
QY 857 KEWV---CNLYKLDNDKIKMLPEAKKYST-----SVKTLSSSSMOPLSLTPODK 904
DB 818 KVMLECEINVISNIDTOYVERIEBAKVLSDSINYIKDEFKLIESTDALCDLKQOE 877
QY 905 PE---VSAND--DTSHTNLNNSLKLFEINISLCKNNIYQELIGOKSSENFYEKILKD 958
DB 878 LEDSHFISPEIDSEIDEGFSIRFINKEGESIFVETEKITFSEY-----ANHTEEISKI 932
QY 959 SDFYNESFTNFVK-----SKADDINSLNDESKRRKLEEDINKLKTLO-----LSFDLYN 1009
DB 933 KCTIDFYNGKLVKRVNLDJTHEVNTLNAAPFIQSLIE--YNSKESLSNLVAMKVOYVA 991
QY 1010 K-YKLEKLERLFPKRRKTVG-----KKMQIKKLT-- 1036
DB 992 QFSTGICLNTITPAKVELYSTALDETDLDLPTLSBGLPIATIIDGVSGLAKIKELSET 1051
QY 1037 ---LLKEQLESKLN-----SLNPKH 1054
DB 1052 SDPLRLQETIEAKIGIMAVNLTTATTAITSSLGIASGFSILVPLAGISAGISLVNNEL 1111
QY 1055 VLQNFVS---FNNKKEKEIETEN--TLENTKILHLK-----YKGLVK 1093
DB 1112 VLDRKATKVDYF---KHVSLVETEGVFPLDDKIMPDODLVISEIDFNNSIYVKGCE 1168
QY 1094 YNNGESSPLKLTSEE-----SIQEDNVASLNFVLSLECKIKONL---NLEKK 1141
DB 1169 IYRMGGSGHTYDIDHFFSAPSTIYREPHLSTIYOLEVQKEEDLSKDLVLPAPNR 1228
QY 1142 KLSY---LSSGLHHLI---AELKEVIKNNKNTGN-----SPSENNTDV 1178
DB 1229 VAWMETGWTPLGRSLLENQGTKLDRIRD--NYGEFYWRFAFIADALITTLKPREDTNII 1287
QY 1179 NNALESYKFLDEGTQVATVWSESGSDTLEOSQPKKPASTHGAASNTITTSQ----- 1231
DB 1288 RINDSNTR---SFIYPII---TTEYIRKLSYFSYSGGTVALSLSQYMMGAINI 1336
QY 1222 -----NVDEVDVYIIPIGSESEEDVDLGOVVTGGAAPPVYIDNLISKIENY 1281
DB 1337 ELSESDWTIIDVNAVROVTTI---ESDKIKKGDLEGLSLTSLIEENKI--TLNSH 1387
QY 1282 EYLVLLKPLAGYRSLSKOLENNVMTFNVNVKDIILMSRFNKRNFKNVLESDLI--PYRDL 1339
DB 1388 EINFSGEVNG-----SNGFVSLTFSILEGINA-----IIEVDLSYSYKLL 1428
QY 1340 TS-----SNYYKD-----PYFLNKEKDKFLSYNTYKDSITDIDI 1376
DB 1429 ISGELKIILMSNHIOOKIDYIGFENSELOKNIPYSFVSEGEN-----GFLNGSTKGL 1483
QY 1377 NFANDVLGYKLLSEKYSDDSI---KKYINDQGENEKEKLPPLNNIETLYKTVNDKI 1432
DB 1484 -----FVSELDVYLVISKVYDDSK---PSFGYSNNLKDVKVYTKDNV 1524
QY 1433 DLFVITHEAKVNLTYEKSNEVKKIKELNLYLKTIDOKLADFKNKNFV--GIADL----- 1485


```

OY 1037 ---LKEQLESKLN-----SIINPKH 1054
||:||||:
Db 1052 SPPILROELEAKIGIMAVNLTTATTAITSSIGIASGFSILVPLAGISAGIPSLVNNBL 1111.
OY 1055 VIQNFV---FENKKEAEIAETEN--TLENTKILKH-----YKGLVK 1093
||:||||:
Db 1112 VLROKATKVVDF---KHVSLVETGVFLDDKIMMPDDLVISEIDENNNISYLGKCE 1168
OY 1094 YNGESSPLKLTSEE---SIOTEDNVASLENFKVLSLECKKLNL---NLEKK 1141
||:||||:
Db 1169 IWRMEGGSGHVYTDIDHFFSAPSITTYRPHLSIYDLEVOKELDLSKDLVLPAPNR 1228
OY 1142 KLSY---LSSGLHLI---AELEKVIKKNKNTGN-----SPSENNFDV 1178
||:||||:
Db 1229 VFAMETGWTPLGRSLJENCTKLLDRD-NYEGEYWRFAFIADALITTLKPREDTVI 1287
OY 1179 NNALESYKRPPEGTAVTVSESGSDTLEOSQPKRPASTHVAESNTITTSQ----- 1231
||:||||:
Db 1288 RINLSNTR---SFIWPII---TTEYIREKLSYFYGSGGTALSLQYNNGINI 1336
OY 1232 ---NYDEVDVYIIPRIGESSEDYDDLQGVYTGAVTPSVYDNLISKIENY 1281
||:||||:
Db 1337 ELSESDVMIIDVNVVAVDTI---ESDKIKKGDLLEGILSTLSEENKI--ILNSH 1387
OY 1282 EYLYLKPLAGYRSLKOLENNVMTFNVVKDILNSRFNRKRNFKNVLESGLI--PYKDL 1339
||:||||:
Db 1388 EINFGEVNG-----SNGFVSLTFSILEGINA-----IIEVDLSISKLL 1428
OY 1340 TS-----SNVVKD-----PYKLNKKRDKFLSYNYIKDSIDTDI 1376
||:||||:
Db 1429 ISGELKILMLNSNHIOQKIDYIGFNSLOKNIPYSFVDECKEN---GFIINGSTREGL 1483
OY 1377 NFANVNLGYKRLSKRYKSDSI---KKYINDQGENEKLPLPNIETLYKYVNDKI 1432
||:||||:
Db 1484 ---FVSELPDVLVLSIKVYMDSK---PSRGYSNNMLKDVKITTKDMV 1524
OY 1433 DLFIHLBAKVLNTYKESNVEVKIKELNLYKTIODKLADFKNNNFV---GIADL--- 1485
||:||||:
Db 1525 N-----ILTYGLKDDIKISIS-----LTLODE-KTILNSVHLDSESVAEILKWM 1569
OY 1486 ---STUDYNNHNLTKFLST---GWFEENLAKTVLSMLDGN--LQGLMLNISQHCYKQCP 1538
||:||||:
Db 1570 NRKGTNTSDSLMSPLESMINKISIFVNFLOSNIKFTLDANFIISGTSIGQEFPI---CD 1626
OY 1539 QNSGC-----FRHL-----DERECKCLNRYKQEG---DK 1565
||:||||:
Db 1637 ENDNIQPYFIKFNLTETNYTLYVGNROMKIYEPNVDLDDSGDISSTVINFOSKLYLGIDS 1686
OY 1566 CVE---NPNPTCNENN 1578
||:||||:
Db 1687 CVNKVVISPNITYTDEIN 1703

```

```

RESULT 51
US-08-915-136-10
: Sequence 10, Application US/08915136
: Patent No. 6290960
: GENERAL INFORMATION:
: APPLICANT: KINK, JOHN A.
: APPLICANT: THALEY, BRUCE S.
: APPLICANT: PADHYE, NISHA V.
: APPLICANT: FIRCA, JOSEPH R.
: APPLICANT: STAPFORD, DOUGLAS C.
: TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
: PREVENTION OF C. DIFFICILE DISEASE
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA

```

```

: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/915,136
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/480,604
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/405,496
: FILING DATE: 16-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/329,154
: FILING DATE: 25-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,907
: FILING DATE: 02-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/985,321
: FILING DATE: 04-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/429,791
: FILING DATE: 31-OCT-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: INGOLIA, DIANE E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: OPND-01763
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 397-8338
: TELEFAX: (415) 705-8410
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2366 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-915-136-10

```

Query Match 3.3%; Score 278.5; DB 4; Length 2366;
 Best Local Similarity 18.5%; Pred. No. 1.2e-06;
 Matches 347; Conservative 292; Mismatches 595; Indels 643; Gaps 92;

```

OY 121 DAKSYADL-----KHRVRYLL-TIKELKYPOLFDLTNMLTLC-D-NHGFXYL----- 167
||:||||:
Db 51 DINSLDIYIDTYKKSGRKALKKRFREYLVTEVLELKNNNLTPVEKNLH-FWVIGGQIND 109
OY 168 -----IDGTEELNELLYKINFFD-----LLRAKINDVCANDYCOQIPFNLIKRIANEID 215
||:||||:
Db 110 TAINIYNQKDVNS-DYNNVNFYDSNAFLINTLKKTVESAINDTLE-SFRENLDPRED 167
OY 216 VLKTLVFGYRKPLDNIKQNVGKMEYIKKNK-----KTIEINE 254
||:||||:
Db 168 YNK-----FFRKMEITIIYKQKNFINTYKAQRENEBELIIDIVKYTILSNESKEIDELNT 223
OY 255 LIEESKTIIDKKN-NATKEERKKLYQAQYDLSIYNKOLEAHNLISYLE-KRIDTLKKN 312
||:||||:
Db 224 YIEESLNTKTONSGNDVNRFEFEKNGES---FNLYEQELVERMNLAAASDLIRISALKEI 280
OY 313 ENI-----KELL-----DKINEINPPRANGNTPNLLDKKNKIEEH-----EK 352
||:||||:
Db 281 GGMVLDVDMPLGIQDPLEESTIEKPSVYVDEWEMTKLEAIMKYKEYIPEYTSHEFMDLME 340
OY 353 EIKE-----IAKTIKFINDSLFTD-----PLELEYLRKKNKNINISAKVETKESTEPNE 402
||:||||:
Db 341 EVQSSPESYLASKDSKSELFSSLGMEASPLEYK--IAFNSGIIINOGLISYKDSYCSUL 398
OY 403 YPNGV--TYPLSYNDINNALNELNASFGDLINFDYTKPSKNIYTDNERKKFINIKE-- 458
||:||||:

```

Db 399 IYKQIENRYKLLNNSLNPASEDNENFTNTNFIDISIMEAN--ADNGR--FMMELGKYL 454
QY 459 -----KRIEKKIESDOKKSYEDRSKSLNITKEYEKLEMBEYDOKRNNNDLTFN- 509
Db 455 RUGFPDVKTTILSGPEAAAYODL-----MFKEGSNHILHEADLNFEISKTNIS 509
QY 510 ---EKMGRKRYSYKVEKLTJHNTFASYENSKHNLKLTALKYME-----DYSLRN 557
Db 510 OSTEQEWASLMSFPDARAK-----AOFEERYKR-----YFEGSLGDDNLDGS-QN 554
QY 558 IYVEKELKYKKNLSIKIENIET-----LVENIKKD-----EOLPEKRI 597
Db 555 IYVDKE--YLEKISLSARSSERKGYIHYIYOQGDKISYEACNLAKPTPYDVLFOKNI 612
QY 598 TKDE-----NKPEKILE-----VSDIVKVOQV-----LML 625
Db 613 EDEIAYVYVNPBGGEIOEDKYKIPSLISDRPKIKLFFIGHGDEFTDIFAGDVSLS 672
QY 626 NKTD---ELKKTOLLKNEVLKHNHVPNSYKOE--NKQEPYLYLVKKEIDKLKVPKY 681
Db 673 TEIEAIDIAKEDISPKSIEI--NLGCMFYSYSINVEETYPGKLLKVKDKISELPSI 730
QY 682 --BSLI--NEEKKNIKTEQSDNSEPTEGELTGOATKPGQAGSALGDSVQAQOE 736
Db 731 SODSIIVASAOYEVRIINSEGRRELDHSGEW----- 761
QY 737 OKOQOPVPVPEAKAOVPTPPAPVNNKTEVNSKLDYLEKLYEFLNTSYCHRYLVSH 796
Db 762 -----INKESIJK-DISSKEYISNFP----- 782
QY 797 STYNEKILKOYKITEBESKSLSCDPLDLLFNIONNIPWYSMFDLSNLSOLEMIYE 856
Db 783 -----KKNKITVRSK-----NLPELSTLQOEIRNNSNSDIDLEE 817
QY 857 KENV-----CNLYLKNDKDKIKNLEBAKRYST-----SVTLLSSSQPLSLTPQDK 904
Db 818 KVLTECEIWNISNIDQIYEEKREKKNLTSDSINYIKOEKFLIESIADCDLKOONE 877
QY 905 PE---VSAND--DTSHSTLNNSLKLEFENILSGKNKNIYOLIGOSSENEFEKILKD 958
Db 878 LEDSHFISFEDISETDGBGFIREFINKETGESIEVETEKTFSEY-----ANHTEELSKI 932
QY 959 SDFEYNESTFNKYK---SKADDINSUNDESKRKKLEEDINKLKTUO-----LSFPLYN 1009
Db 933 KGTIFPTVNGKLVKKNLDTTJHEVNTLNAAFIOSLIE--YNSSKESLSNLSVAMKOVYA 991
QY 1010 K-KYKLEERLFDKKKTGY-----KXKMQIKKLT-- 1036
Db 992 QLFSTGLNTTDDAKAVELVSTALDETIDLLPPLSEGLPIIATIIDGVSIGAAIKELSET 1051
QY 1037 ---LKEQLESKLN-----SLNPKH 1054
Db 1052 SDPLLRQELIEMAGVAVNLTATATITSSLGASGFSILVPLAGISGIPSLVNNEL 1111
QY 1055 VLONFSV---FNNKKKEAIIAETEN--TLENKILILK-----YKGLV 1093
Db 1112 VLDRKATKVVDF--KHVSLVETEGVFTLLDDKIMMPQDVLVISEIDFNNNSIVLGKCE 1168
QY 1094 YVNGESSPLKTLSEE-----SIOEDNYVASLENKVLKLEKGLKDLU-----NLEKK 1141
Db 1169 IMWEGSGHVTVDIDHFFSAPSTIYREPLHSIYDYLEVQKEBLDSLKDLMLVLPNPNR 1228
QY 1142 KLSY---LSSGLHHLI---AELKEVINKNVYTG-----SPENNTDV 1178
Db 1229 VFAMEIGMTGLSLSDGDKLDRIRD--NYEGEFYRYRFAFTADALITLTKPRYEOTNI 1287
QY 1179 NNALESYKKELPBGTDVATVYVSESGDTLEQSPKRPASTHVGAESNTITTSQ----- 1231
Db 1288 RIMLDSWTR-----SFLVPII-----TTEYIREKLSYSFYSGGYVALSLSQYMGMINI 1336
QY 1232 -----NVDEVDVVIIVPIFGESBEDYDLDGOVATVGEAVTPSVIDINILSKINEY 1281
Db 1337 ELSSEDMWIIDVNVADVTI-----ESDKIKKGLDIGILSTLSIEENKI--ILNSH 1387

QY 1282 EVLYLPLAGVYRSKLLKQLENNVMTFVNVKDLILNSKFNKRENEKNVLESOLI--PYKDL 1339
Db 1388 EIMFSGEVNG-----SNGFVSLFSLLEGINA-----IIEVDLLSKSVKLL 1428
QY 1340 TS-----SNVVKD-----PYKLNKROKDFLSSVYITDSDITDI 1376
Db 1429 ISGELKILMNSHIOCKIDYIGFNSLOKNIPIYSPFDSGKEN-----GFIGSTKEGL 1483
QY 1377 NFANDVLGYKKILSEKYSKSLDSI-----KKYINDKOGENEKYLPLNNIETLYKTVNDKI 1432
Db 1484 -----FVSELPRVYLISKYIMDSK---PSFGYSNNLKDVKVYTKDNV 1524
QY 1433 DLEVIHLEAKVLYTEKSYNEVWKIKELNYLKTQDKLADFKKNNFV---GIADL--- 1485
Db 1525 N-----ILNGYVLKDDIKISLS-----LTLQDE-KTIKLNSVHLDSEGAELIKFM 1569
QY 1486 --STDYHNNLLKFLST--GAVFEFLAKTVLSNLLDGN--LQGMNISOHCYKQCP 1538
Db 1570 NRGNTFTSDLSFLESNMKISIPVNFLOSNIKFIIDANFTISGTSIQOFEPT--CD 1626
QY 1539 QNSGC-----FRHL-----DEREDCKLLNYKQEG---DK 1565
Db 1627 ENDNIOYFIKFTLENTLYLYGNRQNMIVERNYDLDDSGDSSIVINSQKYLIGIDS 1686
QY 1566 CVE---NPNPTCENN 1578
Db 1687 CVNKVVISPNITYDEIN 1703

RESULT 52
US-08-957-310-10
; Sequence 10, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-Oct-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-Oct-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-Dec-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-Dec-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-Oct-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPMD-01121
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-310-10

Query Match 3.38; Score 278.5; DB 4; Length 2366;
Best Local Similarity 18.5%, Pred. No. 1.2e-06;
Matches 347; Conservative 292; Mismatches 595; Indels 643; Gaps 92.

QY 121 DAKSADL-----KRVNRYLL-TIKELKYPQLPDLTHMLTCD-NHGRKYL----- 167
DB 51 DINSLTIDYIDYKSGSKKALKRKEKLYTEVELKNNNTLPVKNKH-FWIGIQIND 109
QY 168 -----IDGEEINELLYKLNFFD-----LLRAKLVDCANDYCOIPLNKKIRANEID 215
DB 110 TAININOMKDVNS-DYVNVNFDYDANFLINTLKTIVESAINDTLE-SFRENLDPRFD 167
QY 216 VLKRLVGRKPLDNKDNVGMEDYIKRK-----KTIENINE 254
DB 168 YNK-----FFRRKMEIYDKQKNFIYKKAORENPELIIDIVKTYLSNEYSKEIDELNT 223
QY 255 LIESKRTIDNK-NATKEEKKKLYQAOYDLSYNNKOLEAHNLISVLE-KRIDTLKN 312
DB 224 YIEESLKNITONGNDVNFEEFKGES---FNLYEDELVERMNLAAASDLIRISALKEI 280
QY 313 ENI-----DKINEIKNPPANSNTPTLLDNKKKEEH-----EK 352
DB 281 GGMVLDVDMLGIDPDLFESIEKPSVTVDFEMWTKLEAIKKYKYEITYSEHMDIDE 340
QY 353 EIKE-----IAKTIKYNISLFTD---PLELEYLREKKNKNDISAKVETKESTEPNE 402
DB 341 EVOSFESVLAISKDSKSEIFSLGDMEASPLEVK--IAFNSGIIINOGLISVKDSCSNL 398
QY 403 YPNGV--TYPLSYNDINNALNELNFEGLINPDEYTKEPSKIYTDNRRKKFINIKE-- 458
DB 399 IVKOIENRYKLLNSLNLAISEDNFTTNTTFIDISIMAEAN--ADNGR--FMMLGLGYL 454
QY 459 -----KIRIEKKKIESDKSYEDRSKSLINDITREYEKLEINETYDKFNNNIDLTFF- 509
DB 455 RVGFPPDVKTITINLSGPEAYAAAYODL---MEKESGMNHLLEADLRNEISKTNIS 509
QY 510 ---EKMGKRYSYKVEKLTHTNTFASYSNKHLEKTLKALYME-----DYSLRN 557
DB 510 QSTDEMASLWSPDARAK-----AOFEYKRN-----YFEGSLGEDDNIDFES-QN 554
QY 558 IVEKELKYYNLSKINEIET---LVENIKRD-----EEOLEFEKI 597
DB 555 IYVQKE--YLEKISSLARSSERGIYHYIYVLOGDKISYEAACNLFAKTPYDSVYLQKNI 612
QY 598 TKDE---NKPDEKILE-----VSDIYKVOYQV-----LLM 625
DB 613 EDSEIAYYNGDGEIQEIDKYKIPSIISDRPKIKLTFIGKQDEFNDFIFAGFVDSLS 672
QY 626 NKID---ELKQTQLKANVELKHNTHVNSYKQF-NKOEPIYLLVLAKEIDKALFKPMKV 681
DB 673 TEIEAIDLAKLEDISPKSIEI--NLAGCMNFSYSINVEETYPGKLLLVKKKISELMSSI 730
QY 682 --ESLI---NEKKNIKIRGOSDNSEPTSEGITQATTKPQOQGSALGDSVQAQOE 736
DB 731 SODSIIYSANQYEVAINESGRRELLDHSGE-- 761
QY 737 QKQAOQPVVPEVPEAKAQPPTPPAPVNNKTENVSKLDYLEKLYEPLANTSYICHKYLIVSH 796
DB 762 -----INKESIIK-DISSKEYISFNP----- 782
QY 797 STMNEKILKQYKITEESKISSCDPLDLFRIONNINIVMTSMFSLNNSLSQLEMEIYE 856
DB 783 -----KENKITVYSK-----NPELSTLLLOEIRNNSSSDIELEE 817

QY 857 KEMV-----CNLYKLKDNKIRKMLEEAKKVS-----SVKTLSSSQMPLSLTPDOK 904
DB 818 KVMLECEIINVLSNIDTOIVERIEEAKNLTSDSINYIKDEKFLIESIDALCDLKQONE 877
QY 905 PE---VSAND--DTSHTNLSNLSKLEFNILSLCKNNIYQELIGKSSSENFYEIKLD 958
DB 878 LEDSHFISFEDSEIDSEGFSEIFINKETGEISFVETERTIPSEY-----ANITFEISKI 932
QY 959 SDTFYNESFTNFKV-----SKADDINSLDESKRKKLEEDINKKTKLO-----LSFDLYN 1009
DB 933 KGTIDPTVNGKLVKKNVDLTHTVNTLNAAFIQSLIE-YNSKSELSNLSVAMKVQYVA 991
QY 1010 K-YKAKLERLPDKKTTVG-----KYMQIKKLF-- 1036
DB 992 QLESTGLMTITDAARVELVSTALDETIDLLPTLSEGLPIATIIDGVSLGAIKLESET 1051
QY 1037 ---LKEOLESKLN-----SLNPKH 1054
DB 1052 SDPLLRQLEAKIGIMAVNLTATTAITTSIGIASGFSILLVPLAGISAGIPLSVNNEL 1111
QY 1055 VLQNFV---FENKKKAELIAETEN--TLENTKILKH-----YKGLVK 1093
DB 1112 VLROKATRVVDYF---KHSVIVETEGVFTLLDDKIMMPQDDLVISEIDFNNSIYLKCE 1168
QY 1094 YNNGESSPLKTLSEE-----STQEDNYASLENFVLSKLEGLKDNL---NLKK 1141
DB 1169 IWRMGSGGHVYTDIDHFFSAPSTIYREPHLSITDYLEVQKEEDLSKDLVLPNAPNR 1228
QY 1142 KLSY---LSSGLHLI---AELKEYIKKNNTGN-----SPSENNTDV 1178
DB 1229 VFAMETGWTGRLSLNENGTOKLDRID- NYEGEYWHYFAPFIADALITTLKPRXEDINI 1287
QY 1179 NNALESYKFLPECTDVAIVVSESGDTLEOSQPKPASTHVGASNTITTSQ----- 1231
DB 1288 RLINDSNTR--SFIVPII--TTEYIREKLSYFSGSGTYALSQYMMGINI 1336
QY 1232 -----NYDDEVVYIIVPIGESEEDVDLGOVATGVAVPYVIDNLSIKENEY 1281
DB 1337 ELSESDWMIIDVANDVRYDTI-----ESDKIKKGDLEIGLSTLSEENKI--ILNSH 1387
QY 1282 EYVLKPLAGYRSLSKOLENNVMTFNVVNDILNSRKNRKNFNVLESDLI--PYRDL 1339
DB 1388 EINFSGEVNG-----SNGFVSLTPSLLEGINA-----ILEVDLSRSKYL 1428
QY 1340 TS-----SNVYKD-----PYFLNKKERDKFLSSYNIYKDSIDTDI 1376
DB 1429 ISGELKILMLNSNHIOQKIDYIGFNSLOKNIPIYSFVDSSEKEN-----GFLNGSTKEGL 1483
QY 1377 NFANDVLGYKILSEKYSDDLSI---KRYINDQGENEKYLPPLNNIETLYKTVNDKI 1432
DB 1484 -----FVSELPDVLISKYVYKDSK---PSRGYSNNLKDVKYVTKRNV 1524
QY 1433 DLFWIHLBAKVLNYYEKSANVEVRIKELNLYKTIODKLADFPKNNPV--GIADL--- 1485
DB 1525 N-----ILTYGXYLKDDIKISLS-----LTLODE-KTIKLVNHHDESGVAIELKFM 1569
QY 1486 ---STDYNNNNLLTKFLST---GAVFENLAKYVLSMLDGN--LOGMLNISOHQCYKKQCP 1538
DB 1570 NRGKMTNDSLSMFLSESMNIRKISIFVNLQSNIKFIIDANFIISGTSIGQEFY---CD 1626
QY 1539 QNSGC-----FRHL-----DERECCGLANYKQEG---DK 1565
DB 1627 ENDNIOPIFKENTLETWTLYVGNRMQMIYEPNVDLDDSGDISSTVINFSQKLYXIGDS 1686
QY 1566 CVE---NPNPTCENN 1578
DB 1687 CVNKVVISPNITYTDEIN 1703

RESULT 53
US-09-336-447A-76
; Sequence 76, Application US/09336447A

Patent No. 6310190
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: AEBI, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACIVER, ISOBEL
APPLICANT: FISKE, MICHAEL J.
APPLICANT: FREDENBURG, ROSS A.
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY:024
CURRENT APPLICATION NUMBER: US/09/336,447A
NUMBER OF SEQ ID NOS: 98
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 3788
TYPE: PRF
ORGANISM: Moraxella catarrhalis
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1036)..(3786)
OTHER INFORMATION: Xaa = any
US-09-336-447A-76

Query Match 3.3%; Score 278.5; DB 4; Length 3788;
Best Local Similarity 17.2%; Pred. No. 2.1e-06;
Matches 307; Conservative 290; Mismatches 689; Indels 495; Gaps 72;

QY 32 LEALEAVLTVGLYLPQKEK-----VLNBTSGTAVTTPPGSGVSAAGSGGSA 83
DB 1258 INLESNVEGLXXLSSEQIDNPRTTEINLEARNAXA-NYATPSTINSALFNGESEQ 1316
QY 84 SGG-----SVASGGSVAGSGSVASGSGSRTNPSSSDSAKSYA----- 126
DB 1317 IDNPRTTEINLEARNAXASTDGLKXVSEQIDNPRTTEINLEARNAXANATPSTINSI 1376
QY 127 -DKHVRNLTLEIKELKTPQL-FDLTN-----MLTLDNIGCFKLIDGYE- 172
DB 1377 TDLGTIVDFEXXSEQIDNPRTTEINLEARNAXANATPSTINSITDLGTIVDSEQI 1436
QY 173 -----EINELLYKLFYFDLRAKLN-----DVCANDYCOQPFMLKIRANLVDLKLKLVFG 223
DB 1437 DNPRTTEIN-LEARNANVADLTKVNALDKXVSDXTSEQIDNPRTTEINLEARNAXAN 1495
QY 224 YRKLPLDNKDNVGMEDYIKKKNKTENINELIEESK---KITDKKNATKEE---EKK 276
DB 1496 YATPSTINSALAEQ-----LNGKTLTPVSEQIDNPRTTEINLEARNAXAKHDASTEGK 1551
QY 277 KLYQAOYDLSYKOLE-----EAHN--LISYLEKRIOTLKKNENIKELDKINETKNP 329
DB 1552 KMSEQID-----NPRTTEINLEARNAXALESNVEEGLDLSGSEQIDNPRTTEINLEARN 1607
QY 330 PANGMTPTLDDKKNKIEHEKEIKEIKTIFNIDSLFTDPLELEYLREKKNNDIS 389
DB 1608 NANO-----HTLIKTANKSE-----QID-----NPRTTEIN 1633
QY 390 AKYETESTEPNYPNGVYPLSYDNNALNELNFGDLINPDYTKESKNITYDNR 449
DB 1634 LINEARNALDKNE-----YSKSEQIDNPRTTEINLEARNNA-SITDLGTSKSEQIDNPR 1866
QY 450 KR--FINEIK--EKIKIEK-KKIESDKSY---EDRSLSLDITKEYKLLINEITYDK 499
DB 1687 TEINLEARNANQNTLIESEQIDNPRTTEINLEARNALAE--QOLETTLKSEQIDN 1744
QY 500 FNNNIDLTFNEKMGKRYKVE-----KLHNTFASVENSKNHLEKTLKALKYMEDY 553
DB 1745 PRTEINLEARNANSSDSQIDNPRTTEINLEARNANADADASETLTKSEQIDNPR 1804
QY 554 SLRNIYVERELKYKLNLSIENIEFIVENIKDESOLEFEKRTDKENPKDKITLSEVSD 613
DB 1805 TEINLEARNANAFATAIKDKSE---QIDNPRTTEINLEARNANAKASSNTQNIKASEQ 1861
QY 614 IYKVOYOKVLLMKNIDELKKTQLILKNVELKHNHVPNSYKQENKQEPYLLIVLKREIDK 673

DB 1862 IDNPRTTEINLEARNARLLDQSEQIDNPRTTEINLEARNANAAATADATNGXS 1917
QY 674 LKVFMPKVE-SLINEERKNK-TEGOSDNSE-----DSTGEIRGOATTKGQAGSAGE 727
DB 1918 EQIDNPRTTEINLEARNANAKAXANDRESEQIDNPRTTEINLEARNANANATPSTISI 1977
QY 728 DSVQAO-AOEQOQAPVPVPEAKAQPPTAPVYN--KNTENVSKLDTLEKLYEPLNT 784
DB 1978 NSNQAIDIAQNOTDIODLAAVNELOS--QIDNPRTTEINLEARNANQAD----- 2025
QY 785 SYCHKIYLVSHSTMEKILKYKIKTEESKLSQDPLDLFNIGNNIPVMSMPSDLN 844
DB 2026 -----IANNT----- 2030
QY 845 NSLSQFMETIEKEKVCNLYKLNKNDKIK-NLEAKKYSTSVKTLSSSSMQPLSTPOD 903
DB 2031 NNITELAQOQDQSEQI-----DNPRTTEINLEARNANVNEQOT--EALDALNSEQID 2080
QY 904 KPEVSAN-----DTSHTNLSNLSKLEPENILSLGKNKNYIOELIGKSSENFYE 953
DB 2081 NPRTTEINLEARNALIGDAIVNSQDS-----EQIDNPRTTEINLEARNAKALSNEVE 2136
QY 954 KILKSDTFINESFTNFWVSKADINSLNDESKRKLEEDINKLTKLOLSPDLNKKYL 1013
DB 2137 EGLIDL-----SGRSQIDNPRTTEINLEARNALSNVE--EGLLELS----- 2179
QY 1014 KLERLFDKKTGVKYMOKIKLFLKEQLESKLSLN-----NPKHVLNFSVFFN--KK 1066
DB 2180 -----GRTIQRSEQI-----DNPRTTEINLEARNANQAHANNINNYIELAQ 2224
QY 1067 KEAEIAETENTLENTILKLVKGLVKKYNGESSPLTKTISESIQTEDNYASLENFKVLS 1126
DB 2225 ODOSEQIDNPRTTEINLEARNAXANY-----ATPSTINQA-DIAQNOTDIODLAAVN 2279
QY 1127 KLEGRKLKN-----LNEKKKLSLSSGLHLLINLEKEVYKKNKYTGNSPENNTDVNNAL 1182
DB 2280 ELQSEQIDNPRTTEINLEARNATHDYNERQTEASEQIDNP-----RTEINLIN 2329
QY 1183 ESYKKELEPSTGDAVAVSEGSDTLEOSQPKKASTHVGAESNTITTSQNVDEVDVII 1242
DB 2330 EARN-----AKASSENTONIAKSEQIDNRT-----EINLEARNNA----- 2366
QY 1243 VPIFGSEEDYDGLGOVYTGAVTPSYVIDNLSKINENEYVLYLKLPLAGYRSKLQLEN 1302
DB 2367 -----MILQDT--AIVSN--SODNKTOLEKYK-----SEQIDN 2395
QY 1303 NWTFNWNVNDILNSRFNKNREKNVLESOLIPKDLTSSNYVVKPYKRLNKKRDKPL 1362
DB 2396 P-----RTEINLEARNANAGDTIPLDDXAPSEQIDNP-----R 2431
QY 1363 SSYNYIKDSIDTDINFANDLGYKILSEKY---KSD-LDSIKRYINDKGENEKYLPFL 1418
DB 2432 TEINLEARNANANATPSTINSILHEQOLKGSQIDNPRTTEIN-----LINEA 2483
QY 1419 NNITETLYKTYNDKIDLFVHLEAKVLNITYEKSNNVEYKIKELNYLTKTIDKLADFK-- 1475
DB 2484 RNAXANVATPSTINIF-----FNXSEQIDNPRTTEINLEARNAXANVATPSTI 2533
QY 1476 --NNNFVGADLSTDVNNHNLTKPLSTGMVFENLAKTVSNLLDGLQGLMISQHCY 1533
DB 2534 TINNNNNINNYELAQOQDQSHSDIKTISEQI--DNPRTTEINLEARNAGTGACAGTCA 2590
QY 1534 KQCPQNSGCFRLHDE-----REB--CKC-----LNTY 1559
DB 2591 GATCAGTGACSEQIDNDNALINERDBLBECGACCAACGACGACSEQIDNDNALINE 2650
QY 1560 KQF-----GDKCVENPNTCNEN-----GGC 1581
DB 2651 ARDBLEAGCGGTCCGTGATGACGSEQIDNDNALINERDBLBCTGATCAAGCAGCG 2710
QY 1582 DADAKCTEEDSGSGKKIT-----CECT 1604

Db 2711 GACCGCTSEODJDNALINEARDBLECAAGATCTGCCGCT 27511

RESULT 54

US-09-315-793-52

Sequence 52, Application US/09315793

Patent No. 6221597

GENERAL INFORMATION:

APPLICANT: Roberts, Christopher J.

TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL

TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION

FILE REFERENCE: 9301-048

CURRENT APPLICATION NUMBER: US/09/315,793

CURRENT FILING DATE: 1999-05-21

NUMBER OF SEQ ID NOS: 62

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 52

LENGTH: 1093

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-09-315-793-52

Query Match 3.3%, Score 275.5; DB 4; Length 1093;

Best Local Similarity 18.0%; Pred. No. 6.5e-07;

Matches 216; Conservative 204; Mismatches 409; Indels 371; Gaps 50;

Qy 237 KMEDYK-----KMKKTIENTNELIEESKTTIDKKNKTKTEEEKKLYQAOYD 284

Db 98 KVEDIKGQDVSKIEITLKNSPNVTDI-EYIDARDETIKTRITRSKRSYDLYINDYO 156

Qy 285 LS-----LYNKQLEAHNLISVLEKRIIDTLKKNENIKELDKINEIKRPPANSGN 335

Db 157 VSESVEKTLVQNLQDNLQPLS--QERVEEFARLKVSLVETISI-----204

Qy 336 TPNTLLDKKKIEHEKEIKETIKTIFKNIDSLFTDPLELEYLREKKNKNDISAKVETK 395

Db 205 DASLDVDELRELQNGDSLOKDLDFKAKI-----VHLRQESDKL-----R 247

Qy 396 ESTEENEPNGTYPLSYNDINNALNELNSFGDLNPFDTYKEPSKNITTYDNEKKFFINE 455

Db 248 KSV-----SLRDPONKKGELHLSQL-PY-----272

Qy 456 IKEKIKIEKKIESDKSYEDRSKSLNDITKEYEKLEINEIYDSKFNNDIDLTNEKMMKG 515

Db 273 --VKYKHKEKLNLYKEEYERAKANLRAILKDKRPFAN-----TKK 311

Qy 516 RYKRYVEKLTJHNTFASTENSKHNLKLTALKTMEYSLRNIVEK--ELKYKYNLSK 573

Db 312 TLENOVEELTEKSLKTEFLKAK-EKINEIFEKLN--TIRDEVIKKKNQNEYRGRTKK 368

Qy 574 IENEIETLVENIKKDEQLFEKTKDKENKDPKLEVSIDYKVQVAVLLMKKIDELK 633

Db 369 LQAT-----ISTKEDLRSEOLLAQTHLPKSVFEDIDIKRKET-----INKGELRD 417

Qy 634 TOLILKNELKHNHVPNSYKQENKQEPYLLVLKKEIDKLKVFMPKVESLINEEKKNIK 693

Db 418 ---LISEIDAK-----AINHEMRISIQ 437

Qy 694 TEGOSDNSEPTGEITGOATTKPGQAGSALSGDSVQAQAOEKQAPVPV--PVPEA 751

Db 438 RQAESKTKSLTTTDKI-----GILNODDLKEVDVAVLMEHPEM 478

Qy 752 KAQVTPAPANNKTEVSKIDYLEKLEFLNTSYTCHKIYV-----SHSTMEKILYOY 807

Db 479 KDKILEPPI-MVSAINQAFAYLAQCVY-NTS---KALTVDSDSYKLFANILDKF 532

Qy 808 KITREESKLSGCDPLDLFNQNNIPYMSWFDLSNLSLOLFWIEIEKEKVCVLYLK 867

Db 533 KVNLRLEISSATTPP-----VPA-----ETVRDLGEGGLSDFITG-----568

Qy 868 DNDKIKNLLEBAKAVST--SVKTLSSSMQPLSLTPQDKPEVSNANDJTSHTNLNSLK 924

US-09-315-793-52

Db 569 -DKRYMKMLCQTSKHTIIPVSRRELTPAQIKKL-ITPRPNKTI-----609

Qy 925 LFNELSLGKKNKIYQELIGQ-----SENFYEKILKDSPTFNESTFNPKSKA 975

Db 610 LFKRIHGNRLVDIKQASVGSQVPPDVYSIKQNFYO-----647

Qy 976 DDINSLNDESKRKLLEEDINKLKTQLQSLFDLYNKKLELEFPDKKTKGKYKQIKKL 1035

Db 648 ---GSMNEQKIRIENELINKNE-----YNRKSTLDALSNQK---SGYRHEL---691

Qy 1036 TLKBOLESKLNLSNPKHVLQNFVFPNKKKEAEIAETENTLENTKILKHYKGLVRY 1095

Db 692 ---SELASKNDINREAHQUNEIRKKYTMK-----STIETLRKLDQKREAR--737

Qy 1096 NCESSPLTSPESITQEDNVA--LENFVYSKLEGLKQNLNEKKLS--YLSGLH 1151

Db 738 -----KDVOSKIKDIDQIQOLLKORHLHLSKASSKSLKQCKELISTOILLQPEAQ 790

Qy 1152 HLIALKEVINKKNKYTGNSPSENNTDVNNALESYKKFLPEGTD-----1194

Db 791 NMDVSMNDVIGFPN-----ERADLKSQYEDKKKKYKEMKDPPEQSMREITSYDQ 843

Qy 1195 -----VATVVSSEGS-----DTLEQSQPKKPASTHVGASNTTTSQNVDEVDV 1240

Db 844 TKELINKVAEKEYEEBGNFNLSPVDVLDKLESEIAMVNH---DESAVTILDQVTAELREL 900

Qy 1241 -IIVPIGSEEDYDDL-GQVVTGAVTPSYVDNLSTKIENTEYVLYLK-PLAGYRSLK 1297

Db 901 EHTVP---QOSKDLKETIAKLEHDAVLEPKDIDVSKISARFARLFNNGVSGAGVLR-LE 956

Qy 1298 KOLENNWTFNVNVKDIINSRPNREN-----FNKVESLDIPYKDLTSSNYVKD 1348

Db 957 KPKDYAEKIELMVKFRNAPLAKIDSHTSQSGERAVSTVLR--MIALQEFISAFRYYVD 1014

RESULT 55

US-08-290-919-1

Sequence 1, Application US/08290919

Patent No. 5720959

GENERAL INFORMATION:

APPLICANT: HOLDER, ANTHONY A.

APPLICANT: BLACKMAN, MICHAEL J.

TITLE OF INVENTION: CHAPPEL, JONATHAN A.

TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,919

FILING DATE: 04-OCT-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9203821.5

FILING DATE: 22-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/00367

FILING DATE: 22-FEB-1993

ATTORNEY/AGENT INFORMATION:

NAME: KOKULIS, PAUL N.

REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELE: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label=X
OTHER INFORMATION: /note="X = M and N, or N"
US-08-290-919-1

Query Match 3.3%; Score 274; DB 1; Length 48;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1527 ISOHCVKRCPCPNSGCFRHLDERECKCLLANKOGDKCVENPPT 1573
DB 2 ISOHCVKRCPCPNSGCFRHLDERECKCLLANKOGDKCVENPPT 48

RESULT 56
US-08-685-871-2
Sequence 2, Application US/08685871
Patent No. 6013499
GENERAL INFORMATION:
APPLICANT: NARUMIYA, Shuh
APPLICANT: IMAMATSU, Akihito
TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,871
FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184102
FILING DATE: 25-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-262553
FILING DATE: 14-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-685-871-2

Query Match 3.3%; Score 274; DB 3; Length 1354;
Best Local Similarity 19.6%; Pred. No. 1e-06;
Matches 239; Conservative 195; Mismatches 389; Indels 394; Gaps 56;
QY 412 SYNDINNALNELNSFGDLINPPFYTKPSKNT---YTDNE---RKKFINEIEKIEK 464
DB 286 TVSKINHNKNSL-TFPD---DNDISKE-AKNLICARLTREVRILGRNGVEIKRHL--- 336
QY 465 KRTESDKSYEDRSKSLNDITREKELNLIYSKNNNDLNF---EKMGRKRSYK 521
DB 337 -FEKNDQAMWETIROTVAEVPD-----LSSDIDISNFDLEDKGEETFP 383
QY 522 EKLTHNT---FASYSKSH-----NLEKLTALRYMEDYSLRNIVKELRYRN 569
DB 384 PKAFVGNQLPFVGFTYSNRRYLSANPDNRRSNAKSLQF-SLQKTIYKLEPQD 442
QY 570 LISKIENETLVENIKKDEQLFEKKIK--DENKPKILEVSDIVAVQVYLLMK 627
DB 443 MOKLDEMEQKCRSTNKLQ-----KIMKELDEGNQRRNLE-STVSQIEKMLQHR 494
QY 628 IDELKTQQLIKVLELHNHIVPNYSKOENKOPPYLLYLKKEIDKLKVFMPKVESLINE 687
DB 495 I-----NEYOKRAED-----NE 507
QY 688 EKKNIKTEGOSDNSEPTGEITGOATTPRGQAGSALGDSVQAQAOQKQAPVPVP 747
DB 508 KRRNV-----EKEVS-----TLKDQLEDLKKV----- 529
QY 748 VPEAKNOVTPPAVPNNKTEVNSKLDYLEKLEFNTSYICHYIIVSSTMEKILKQY 807
DB 530 --SQNSQL-----ANEKLSQLOK--QLEBANDLTFESDTAVRLKSHTEKSKSISQLE 579
QY 808 KITKEESKLSGCDPLDLFNIONNIPVMSMEDSLNSLSQLEFMEIYEKEMVCNLYKL 866
DB 580 SLNRELQER-----NRI-----LENSKSQYDKYQYQALILEARRD 616
QY 867 --KONDKIKNLEAKKAVSTSVKTLSSSQMLSTLPQDKPEVSAADTSHSNLNSLK 924
DB 617 RGHDSSEMIGLQARITSLOEVVHLKH-----NLEKVEGERKQADMDHNSKEKNMLE 670
QY 925 LFNIIISLGKNNIYOELIGOKSENFYEKILKDSOTFYNESPTNPKSKA--DDINSLN 982
DB 671 IDLNT---KLSLQQL--QEVNEH---KVYKARLTDKHQSLF--AKSYAMCEMKKLL 721
QY 983 DESK-RKLEEDINKLKLTL-QLSPDLNKKYKLELDPDKTKVYKMQIKKLTLKE 1040
DB 722 EEREAREKAKENRVYQIEKQSMLDVQL-KSQQKLEHLGNKE--RMEDEVNLTQLE 777
QY 1041 QLESKLSLNNPKHVLQNFVFFNKKKEKAEIATENTLTKTLK-HKGLVYKYNES 1099
DB 778 QESNKRLLQNLQELKT-QAEFADNLKLEKQKQEIINTLEAKRLLEFELAQLTKQYRGNE 836
QY 1100 SPLKTLSEESIQTEEDVYASLENFKVLSKLEGRKQNLNLEKKLSTYSSGLHLLIAELKE 1159
DB 837 GQNR-----ELQDQLEAQ-----YFSTLYKQVAKELKE 865
QY 1160 VIKNNKVTGNSPENNNTDVNNALLESYKFLPECTDVATVVSSES---GSDTLQF-----S 1210
DB 866 EIEEKN-----RENLIKQIQLQNEKETLATQDLAETKAESEQALRGLEEQYELTQ 918
QY 1211 QPKKPASTHVGAESNTITTSQNVDEVVDYIIVPIGSESEDDDLQGVYTGAVPSPVY 1270
DB 919 ESKKAASRN-----KQELTDK----- 934
QY 1271 DNLISKIENEYEVLYLPLAGVYRSKLQLENNVMTFVNVVKDILNSRFNKRNFKNVLE 1330
DB 935 DRTVSLKE-----ANSMFL-----KDI-----ELIKREN----- 959
QY 1331 SDLIPYKDLTSSNYYVVKDPKFLNKKERDKFLSSYNYIKDSIDTD-----INFANDVL 1383
DB 960 -----EELTEKKMKKAAEEKYL--EKEEISNLKAAFEKNINTERTKLTQAVNKLAEIM 1010
QY 1384 GYKILISEKYSKSLDSIKTKIINDKQGENEKYLPFLNNIETLYTVVDKIDLPYIHLKAVY 1443

Db 1011 NRKDEKIDRKKANTODLRK-----KEKENRK-----1040
Qy 1444 LNTYREKSNVAVKIKELNYLTKTIDQKDLADFKNNNFVCIADLSDYNNNNLTKL-----1499
Db 1041 LNOEREFN-OMVYKHOKELMDQAOVLVECAHRELLOMOLASKSDIEOLRAKLDDSD 1099
Qy 1500 STGWFENLAKTVLSNLDGMLNLSOHQCVKQCPONSQCFRHLDEREE-----CK 1554
Db 1100 STVASPESADETGDNLPESRIEGLWLY-----PARGNKRKYGKMKQYVVSXK 1148
Qy 1555 CLNTYKQEGDKVCNPN 1571
Db 1149 KILFYNDQDKKESQNSPS 1165

RESULT 57
US-08-755-587-185
Sequence 185, Application US/087555587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
TITLE OF INVENTION: susceptibility gene and uses thereof.
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-755-587-185

Query Match 3.28; Score 272; DB 3; Length 1535;
Best Local Similarity 19.28; Pred. No. 1.5e-06;
Matches 255; Conservative 223; Mismatches 520; Indels 344; Gaps 62;

Qy 267 KNATREKREKKLYOAOVDLSYNNKQLEAHNLISVLEKRIPTLKKNNIKELDKINEIK 326
Db 15 KEACNNKEXQLFIT-----PEADSLSPLOEQYENDKRSKRVSDIKEEVLAA 63
Qy 327 NPPANSQNPNTLLDKNKK-IEEHEKE-----IKEIAKTIKFI-----DSL 368

Db 64 HHPVOHSAEYSDTDFOSOKSPLYDHENASTLLIPTSKDVLNLMISGKKSKYMSDEL 123
Qy 369 FTDPLELEYLRK-----KNIDISAKVETKESTE-----TNEYNGV--TYPLSYNDINNAL 420
Db 124 KNNYESDFELTKNIPKERNDOVCALSPNSKIVELLPEPEKYIRGASPSKVOFNQTNLRV 183
Qy 421 NELNSFGDLINFEYDTKEPSKNIYTDNERKFFINEIKK-----IKIEKKIESD-----470
Db 184 IQKODEETLLPKITVHHDSEELFSDNENNNVFOVANERNRNALONTLEHETDILCYNE 243
Qy 471 ---KKS-----YEDRSKSLNDITKEYEKLNIYD-SKFNNNI-----D 505
Db 244 PIFKNSITVLYADIGDKOATQVSIKKDLS--SNIIYDLAEKKNKSKOHLKTLTGODLPD 302
Qy 506 LTFNEKMMGK-----RYSYKEXLTHHTFASYSN-----SKHLKXTLAKTYME 551
Db 303 ILINDKIPDKIDDCKMAGPLDPISNHSPGSGFTFASNKELIKSEHNKSKMFFKDE 362
Qy 552 DY-----SLRNIYVEKELKYYKNLSKIENEIETLV-----ENIKKDEQLFEKK- 596
Db 363 EQYPTSLACVELVNTLALDNQKLSKPOLINTYSAHLQSSVYVSDCKNHHITPQMSFSKOD 422
Qy 597 ITKDN-KPDEK--ILEVSDIVKVOYVLLMKNIDELKTKQILKVELKHNHVP---650
Db 423 FNSNHLTPSOAKETTELSTLEBSGOF---EFTQPRKPSYILO---KMTFVPENO 474
Qy 651 ---NSTYOENK-----QEPYLIYVLAKEIDKLVMPYVESLINEKKNIKEGSD 699
Db 475 VTILNTTEENDAGLVIMNAPSIOVSSKQEGTVGIKQFAGLAKSDCKNSAGYILD 534
Qy 700 NSEPTGEELITG-----OATTKPGQOA---GSALGSDSVQAOQROKQAPVPVYVPEA 751
Db 535 ENEVFRGFYSAHGKVLNVTALQKAVKLFSDIENISEKTSABVD-----PISLSS 587
Qy 752 KAQVPTPAPVNNKTENVSKLD-YLEKLYEFLNTSYICHKYLVSHTSMNEKILKQYKI- 809
Db 588 K-----FHDSVYSMFKIENHNDKTSKKCOLMLONNIEMTGTGFVEIETENYKIN 638
Qy 810 TKREESKLSSC-----DPLDILFNIONNI-----PYMYSM 839
Db 639 TENEDNKYTAASRNSRNLEFVGSDSSKNDYVCIHKDEKDLPTDORNICLKLSGOLMEG 698
Qy 840 FDSLNNLSLQFMELIYERKEMVNLKLNNDKIKMLLEAKVSTSVK-----TSSS 892
Db 699 NTQIKEGSLDLFLVYKQOETCH-----GNTSNKQOL-TARKTEBNIDOFETDLSPTGA 752
Qy 893 SMQPLSLTPD-----KPEVSANDDTSHSTNLNLSKLFENILSLGKNNIYOEL 942
Db 753 SGKNIYVTKESFNKIVNFFDPKPE-----ELHNSLSKLH---SDIRKNK---MDI 798
Qy 943 IGOKSENFYKELIKDSTFTFNESFTNFVSKADINDLNDESKRKKLEEDINKLKTIO 1002
Db 799 LSHETDTVKKKLIKESVPGTGNOLVTF-----QERPQODEEL---KEPTL 841
Qy 1003 LSFDIYNNYKIKL-----ERLPDK-----KTVGYKMKQIKKLTPLK 1039
Db 842 LGFHTASGKKYKIKTESLDYKKNLEFDEKEGTSSETTSHOWAKTL-KYREACQDLIELAC 900
Qy 1040 EQLE-----SKLNSLNPRHVLQNFVFPNNKKKEALIEAETENTLENTLILKHVGL 1091
Db 901 ETVETITAPCKCKEMQNSLNNDKIVCIETVVPKLLSDMLCTQTEMLKTSKSIFLK-----955
Qy 1092 VKYVNGESSPLKLTSEESIQEDNYSLEN-----FKVLSKLEGLKQDLNLEKKLSYLS 1147
Db 956 KVHVENVEKETAKSFTTCYTNOQSPSYIENSALAVYTSCSRCSVSQTSLEAKO--WLR 1013
Qy 1148 SGLHILIAELKEVINKNKNYNSPSENTDVNNALESYKFKPLPBGTVATVYVSESGDPL 1207
Db 1014 EG---IFPDQPERINTADVGNLSLYENSK-NTIAESQKNH-----PSEKQDPTL 1059
Qy 1208 EOSQPKKPASTHVGAESFTTTTSONVDEVDVVIPIFGSEEDYDLDGOVYGEAVTP 1267
Db 1060 NNS-----SMSNSYSYL-----SDEVYSDSGYLSKKNK--LD 1088

[illegible]

```

1      RESULT 58
2      US-08-290-919-2
3      ; Sequence 2, Application US/08290919
4      ; Patent No. 5720959
5      ; GENERAL INFORMATION:
6      ; APPLICANT: HOLDER, ANTHONY A.
7      ; APPLICANT: BLACKMAN, MICHAEL J.
8      ; APPLICANT: CHAPPEL, JONATHAN A.
9      ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
10     ; TITLE OF INVENTION: VACCINE
11     ; NUMBER OF SEQUENCES: 19
12     ; CORRESPONDENCE ADDRESSES:
13     ; ADDRESSEE: CUSHMAN DABRY & CUSHMAN, L.L.P.
14     ; STREET: 1100 NEW YORK AVENUE, N.W.
15     ; CITY: WASHINGTON
16     ; STATE: D.C.
17     ; COUNTRY: U.S.A.
18     ; ZIP: 20005-3918
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/290,919
26     ; FILING DATE: 04-OCT-1994
27     ; CLASSIFICATION: 435
28     ; PRIOR APPLICATION DATA:
29     ; APPLICATION NUMBER: GB 9203821.5
30     ; FILING DATE: 22-FEB-1992
31     ; PRIOR APPLICATION DATA:
32     ; APPLICATION NUMBER: PCT/GB93/00367
33     ; FILING DATE: 22-FEB-1993
34     ; ATTORNEY/AGENT INFORMATION:
35     ; NAME: KOKULIS, PAUL N.
36     ; REGISTRATION NUMBER: 16,773
37     ; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
38     ; TELECOMMUNICATION INFORMATION:
39     ; TELEPHONE: (202) 861-3000
40     ; TELEFAX: (202) 832-0944
41     ; TELEX: 6714627 CUSH
42     ; INFORMATION FOR SEQ ID NO: 2:
43     ; SEQUENCE CHARACTERISTICS:
44     ; LENGTH: 48 amino acids
45     ; TYPE: amino acid
46     ; STRANDEDNESS: single
47     ; TOPOLOGY: linear
48     ; MOLECULE TYPE: protein
49     ; FEATURE:
50     ; NAME/KEY: Modified-site
51     ; LOCATION: 1
52     ; OTHER INFORMATION: /label=X
53     ; OTHER INFORMATION: /note=X = M and N, or N"
54     ; US-08-290-919-2
55     ; Query Match 3.2%, Score 271, DB 1, Length 48;
56     ; Best Local Similarity 97.9%, Pred. No. 2,4e-08;
57     ; Matches 46; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1527 ISOHOCVKKQCPONSGCFRHLDERECKCLLNKQEGDKCVENPNPT 15733
|||||:|||||
Db 2 ISOHOCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPT 48

```

```

RESULT 59
US-08-923-992A-10
: Sequence 10 Application US/08923992A
: Patent No. 6280738
: GENERAL INFORMATION:
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Blake, Milan S.
: TITLE OF INVENTION: No. 6280738- 19a Fc Binding Forms of the Group B
: TITLE OF INVENTION: Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESSES:
: ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM pc compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/923,992A
: FILING DATE: 05-SEP-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/024,707
: FILING DATE: 06-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: REFERENCE/DOCKET NUMBER: 1438, 0140001/RWE
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-923-992A-10

```

[illegible]

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

SUMMARIES

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using: sw model

Run on: March 30, 2003, 22:00:50 ; Search time 9477 Seconds

(without alignments)
15170.187 Million cell updates/sec

Title: US-09-269-874A-2

Perfect score: 4940

Sequence: 1 cgcacgcgataatcatc.....ttcattcattatgacgatg 4940

Scoring table: IDENTITY:NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database :

GenEmbl:.*
1: gb_da:.*
2: gb_bt:.*
3: gb_in:.*
4: gb_cm:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_sts:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vi:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_om:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_sts:.*
28: em_un:.*
29: em_vi:.*
30: em_htg_hum:.*
31: em_htg_inv:.*
32: em_htg_other:.*
33: em_htg_mus:.*
34: em_htg_pln:.*
35: em_htg_rtd:.*
36: em_htg_mam:.*
37: em_htg_vrt:.*
38: em_sy:.*
39: em_htgo_hum:.*
40: em_htgo_mus:.*
41: em_htgo_other:.*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB ID	Description
1	4940	100.0	4940	6 A92451	A92451 Sequence 2
2	4938.4	100.0	4940	12 CVE131294	AJ131294 Cloning v
3	2816.4	57.0	5917	3 PFAMF195	X02919 Plasmodium
4	2816.4	57.0	7038	3 PEGP195A	X15063 Plasmodium
5	2814.4	57.0	4920	6 A92450	A92450 Sequence 1
6	2706.2	54.8	5760	6 A04562	A04562 P. falcipar
7	2706.2	54.8	5760	6 E00656	E00656 cDNA encodi
8	2686.8	54.4	5282	3 PEP190	X03371 P. falciparu
9	1400.6	28.4	5271	3 AF062349	AF062349 Plasmodu
10	1331.2	26.9	5243	3 AF062348	AF062348 Plasmodu
11	1330.8	26.9	5085	3 AF218248	AF218248 Plasmodu
12	1290.4	26.1	5392	3 PEP190G1	X05624 Plasmodium
13	1290.4	26.1	5754	3 PFAMMSAA	M9143 P. falciparu
14	1272.2	25.6	5312	3 PFAMSA1B	Z53327 P. falciparu
15	1267.2	25.7	5276	3 PFAGP195	M37213 P. falciparu
16	1267.2	25.7	5541	3 PEGP195	X03831 P. falciparu
17	1169.6	23.7	1968	3 AF480451	AF480451 Plasmodu
18	910.4	18.4	1636	3 PFAMSP806	D13357 P. falcipar
19	910.4	18.4	1636	3 PFAMSP808	D13358 P. falcipar
20	908.8	18.4	1636	3 PFAMSP828	D13360 P. falcipar
21	908.8	18.4	1636	3 PFAMSP837	D13362 P. falcipar
22	908.8	18.4	1636	3 PFAMSP844	D13363 P. falcipar
23	907.2	18.4	1636	3 PFAMSP844	D13363 P. falcipar
24	907.2	18.4	1636	3 PFAMSP844	D13363 P. falcipar
25	905.6	18.3	1636	3 PFAMSP8222	D13359 P. falcipar
26	784.6	15.9	3395	3 PFAMMSA1	X61930 P. falcipar
27	711.8	14.4	3518	3 PPAF190A	M35727 P. falciparu
28	711.8	14.4	3518	3 PEP190G	Y00087 Plasmodium
29	691	14.0	1203	3 PFAMSP1	M46081 Plasmodium
30	655.8	13.3	1131	3 AF325919	AF325919 Plasmodu
31	623.6	12.6	1065	3 PFASURFPRO	L20092 Plasmodium
32	570.2	11.5	999	3 PFAMSA1AC	M77715 Plasmodium
33	570.2	11.5	999	3 PFAMSA1AC	M77715 Plasmodium
34	570.2	11.5	999	3 PFAMSA1AC	M77715 Plasmodium
35	568.6	11.5	999	3 PFAMSA1AM	M77725 Plasmodium
36	568.6	11.5	999	3 PFAMSA1AF	M77718 Plasmodium
37	556.6	11.3	1133	3 AF061143	AF061143 Plasmodu
38	547.8	11.1	1080	3 AF286876	AF286876 Plasmodu
39	546.4	11.1	1956	3 PFMEZSA1D	Z53329 P. falciparu
40	546.2	11.1	999	3 PFAMSA1AH	M77720 Plasmodu
41	544.8	11.0	1956	3 PFMEZSA1A	Z53326 P. falciparu
42	540.4	10.9	1141	3 AF061141	AF061141 Plasmodu
43	538.8	10.9	1142	3 AF061140	AF061140 Plasmodu
44	535.4	10.8	1959	3 PFMEZSA1C	Z53328 P. falciparu
45	528.2	10.7	1116	3 AF061146	AF061146 Plasmodu
46	527.8	10.7	1109	3 PFHB3A	X52962 Plasmodu
47	527.2	10.7	1150	3 AF061136	AF061136 Plasmodu
48	512	10.4	1158	3 AF061139	AF061139 Plasmodu
49	508.2	10.3	1102	3 AF061147	AF061147 Plasmodu
50	508.2	10.3	1157	3 AF061135	AF061135 Plasmodu
51	502.4	10.2	1123	3 AF061145	AF061145 Plasmodu
52	500.8	10.1	1124	3 AF061144	AF061144 Plasmodu
53	500.2	10.1	1149	3 AF061138	AF061138 Plasmodu
54	499.6	10.1	1167	3 AF061137	AF061137 Plasmodu
55	498.2	10.1	1129	3 AF061133	AF061133 Plasmodu
56	486.4	10.0	1002	3 PFAMSA1AD	M77716 Plasmodu
57	486.4	10.0	1002	3 PFAMSA1AK	M77723 Plasmodu
58	485.2	9.8	1002	3 PFAMSA1AB	M77714 Plasmodu
59	485.2	9.8	1002	3 PFAMSA1AL	M77724 Plasmodu
60	484.6	9.8	1110	3 AF061132	AF061132 Plasmodu
61	483.2	9.8	1121	3 AF061142	AF061142 Plasmodu
62	483	9.8	1111	3 AF061134	AF061134 Plasmodu
63	483	9.8	1112	3 AF061131	AF061131 Plasmodu
64	473.2	9.6	993	3 PFAMSA1AP	M77728 Plasmodu
65	463.2	9.4	984	3 PFAMSA1AJ	M77722 Plasmodu

66	463.2	9.4	984	3	PFAMSA1AN	M77726 Plasmodium	139	246.6	5.0	354	6	A64813	A64813 Sequence 4
67	457	9.3	1102	3	AF061150	AF061150 Plasmodium	140	243.4	4.9	737	6	A01099	A01099 Artificial
68	455.4	9.2	1101	3	AF061149	AF061149 Plasmodium	141	240.2	4.9	668	6	E01429	E01429 Genomic DNA
69	454	9.2	1141	3	AF061120	AF061120 Plasmodium	142	237.2	4.8	639	3	PFAMSA1AW	M77735 Plasmodium
70	440.4	8.9	1133	3	AF061119	AF061119 Plasmodium	143	228.6	4.6	1774	3	AF199407	AF199407 Plasmodium
71	438.4	8.9	1177	3	AF061122	AF061122 Plasmodium	144	227	4.6	1774	3	AF199405	AF199405 Plasmodium
72	437.4	8.9	963	3	PFAMSA1AU	M77733 Plasmodium	145	217.6	4.4	668	6	I05209	I05209 Sequence 5
73	435.6	8.8	1726	3	PFAMSP8352	D13349 P. falcipar	146	214.2	4.3	371	3	AB016617	AB016617 Plasmodium
74	434	8.8	1726	3	PFAMSP835	D13346 P. falcipar	147	212.6	4.3	371	3	AB016616	AB016616 Plasmodium
75	434	8.8	1726	3	PFAMSP838	D13350 P. falcipar	148	212.6	4.3	371	3	AB016615	AB016615 Plasmodium
76	432.4	8.8	1726	3	PFAMSP807	D13344 P. falcipar	149	212	4.3	442	3	PP190B	PP190B Plasmodium
77	432.4	8.8	1726	3	PFAMSP8351	D13348 P. falcipar	150	210.4	4.3	365	3	D86292	D86292 Plasmodium
78	432.4	8.8	1726	3	PFAMSP842	D13352 P. falcipar	151	209.4	4.2	371	3	AB016618	AB016618 Plasmodium
79	430.8	8.7	1726	3	PFAMSP947	D13351 P. falcipar	152	207.8	4.2	371	3	D86293	D86293 Plasmodium
80	429.2	8.7	1726	3	PFAMSP946	D13354 P. falcipar	153	205.6	4.2	1795	3	AF199409	AF199409 Plasmodium
81	428.4	8.7	1114	3	AF061123	AF061123 Plasmodium	154	202.4	4.1	415	3	PFAMSA1F	M32116 P. falciparu
82	427.8	8.7	1729	3	PFAMSP814	D13345 P. falcipar	155	201.8	4.1	342	6	A64801	A64801 Sequence 6
83	427.6	8.7	1726	3	PFAMSP8221	D13347 P. falcipar	156	201.8	4.1	342	6	A64815	A64815 Sequence 6
84	426	8.6	1726	3	PFAMSP843	D13351 P. falcipar	157	201.8	4.1	456	6	A02300	A02300 Synthetic D
85	426	8.6	1729	3	PFAMSPM20	D13343 P. falcipar	158	201.8	4.1	462	6	A02299	A02299 Synthetic D
86	424.6	8.6	1135	3	AF061125	AF061125 Plasmodium	159	201.6	4.1	1801	3	AF199406	AF199406 Plasmodium
87	422.8	8.5	1090	3	PE3D7A	X52963 Plasmodium	160	201.4	4.1	350	3	D86296	D86296 Plasmodium
88	422	8.5	1104	3	AF061151	AF061151 Plasmodium	161	200.8	4.1	1795	3	AF199410	AF199410 Plasmodium
89	420	8.5	1107	3	AF061148	AF061148 Plasmodium	162	199.6	4.0	1801	3	AF199393	AF199393 Plasmodium
90	420	8.5	1107	3	AF061148	AF061148 Plasmodium	163	195.2	4.0	1801	3	AF199401	AF199401 Plasmodium
91	418.2	8.5	1187	3	PFAMSA1AS	M77731 Plasmodium	164	194.8	3.9	291	6	A64810	A64810 Sequence 1
92	413.2	8.4	1002	3	PFAMSA1AT	M77732 Plasmodium	165	194.8	3.9	291	6	A64810	A64810 Sequence 1
93	413.2	8.4	1002	3	PFAMSA1AT	M77732 Plasmodium	166	194.8	3.9	330	6	A64804	A64804 Sequence 9
94	411.4	8.3	5730	3	PFAPV200A	M77733 Plasmodium	167	194.8	3.9	387	6	A64818	A64818 Sequence 9
95	410.6	8.3	1141	3	AF061129	AF061129 Plasmodium	168	194.8	3.9	387	6	A64802	A64802 Sequence 7
96	409	8.3	1173	3	AF061130	AF061130 Plasmodium	169	194.8	3.9	387	6	A64816	A64816 Sequence 7
97	406.8	8.2	1098	3	AF061121	AF061121 Plasmodium	170	194.6	3.9	309	6	AX278285	AX278285 Sequence
98	398	8.1	939	3	PFAMSA1AO	M77727 Plasmodium	171	194.6	3.9	3147	6	AX278286	AX278286 Sequence
99	395.6	8.0	975	3	PFAMSA1AR	M77736 Plasmodium	172	194.6	3.9	3147	6	AX278287	AX278287 Sequence
100	394.6	8.0	966	3	PFAMSA1AX	AF061124 Plasmodium	173	193.2	3.9	1801	3	AF199395	AF199395 Plasmodium
101	393.8	8.0	1203	3	AF061124	M77729 Plasmodium	174	191.6	3.9	1801	3	AF199402	AF199402 Plasmodium
102	392.2	7.9	1005	3	PFAMSA1AQ	M77734 Plasmodium	175	191.6	3.9	1801	3	AF199403	AF199403 Plasmodium
103	392.2	7.9	1005	3	PFAMSA1AV	X75899 P. falciparu	176	191.6	3.9	1801	3	AF199408	AF199408 Plasmodium
104	392	7.9	698	3	PFMSPIA	X78207 P. falciparu	177	191.4	3.9	309	6	AX278288	AX278288 Sequence
105	392	7.9	698	3	PFMSPIREC	M77737 Plasmodium	178	191.4	3.9	3147	6	AX278289	AX278289 Sequence
106	389.8	7.9	966	3	PFAMSA1AY	M14632 Plasmodium	179	191.4	3.9	3147	6	AX278290	AX278290 Sequence
107	386.4	7.8	1030	3	PFAGP185	AF061127 Plasmodium	180	191.4	3.9	3147	6	AX278290	AX278290 Sequence
108	384.4	7.8	1030	3	PFAGP185	AF061127 Plasmodium							
109	382.4	7.7	5181	3	PFAMSA1G	M0807 Plasmodium							
110	380.8	7.7	657	3	PFMSA1FID	Y10598 P. falciparu							
111	373.8	7.6	1277	3	AF061128	AF061128 Plasmodium							
112	360.8	7.3	635	3	PFU09209	U09209 Plasmodium							
113	360	7.3	720	3	AF251345	AF251345 Plasmodium							
114	353.6	7.2	969	3	PFAMMFA	L10380 Plasmodium							
115	337.4	6.8	893	3	PFAMMSAB	M19144 P. falciparu							
116	335.4	6.8	1116	3	PFU20730	U20730 Plasmodium							
117	333.8	6.8	1116	3	AF040569	AF040569 Plasmodium							
118	333.8	6.8	1116	3	PFU20731	U20731 Plasmodium							
119	332.4	6.7	1119	3	PFU20729	U20729 Plasmodium							
120	332.2	6.7	1116	3	PFU20655	U20655 Plasmodium							
121	332.2	6.7	1116	3	PFU20656	U20656 Plasmodium							
122	330.8	6.7	1119	3	PFU20726	U20726 Plasmodium							
123	330.8	6.7	1119	3	PFU20727	U20727 Plasmodium							
124	330.6	6.7	1116	3	PFU20653	U20653 Plasmodium							
125	329.2	6.7	1119	3	PFU20728	U20728 Plasmodium							
126	329.2	6.7	1119	3	PFU20732	U20732 Plasmodium							
127	327.6	6.6	1116	3	AF040567	AF040567 Plasmodium							
128	327.6	6.6	1119	3	PFU20733	U20733 Plasmodium							
129	327.6	6.6	1119	3	AF040568	AF040568 Plasmodium							
130	325.4	6.6	1119	3	PFAMSA1AI	M77721 Plasmodium							
131	325.4	6.6	1119	3	PFAMSA1AI	M77721 Plasmodium							
132	322.8	6.5	1249	12	AF395808	AF395808 Synthetic							
133	308.8	6.3	660	6	A02298	A02298 Synthetic D							
134	308.8	6.3	1068	6	A02301	A02301 Synthetic D							
135	282.6	5.7	639	3	PFAMSA1AE	M77717 Plasmodium							
136	272	5.5	654	3	PFMSA1FIS	Y10599 P. falciparu							
137	268.8	5.4	654	3	PFMSA1FIS	Y10600 P. falciparu							
138	246.6	5.0	354	6	A64799	A64799 Sequence 4							

ALIGNMENTS

RESULT 1
LOCUS A92451 4940 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 2 from Patent WO9814583.
ACCESSION A92451
VERSION A92451.1 GI:6741181
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 4940)
AUTHORS Pan, W. and Bujard, H.
TITLE METHOD FOR PRODUCING RECOMBINANTS INTENDED FOR USE IN A COMPLETE
JOURNAL MALARIA ANTIGENE (p190)/MSPI
PATENT: WO 9814583-A 2 09-APR-1998;
PANTING (DE); BUJARD HERMANN (DE)
FEATURES
source 1..4940
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 1669 a 1157 c 1060 g 1054 t
ORIGIN
Query Match 100.0%; Score 4940; DB 6; Length 4940;
Best Local Similarity 100.0%; Pred No. 0;
Matches 4940; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	CGACGGGTATGAAATCATTTTCTTCCTCTGTCACTTCTGTTTTTATCATCAATCT	60
Db	1	CGACGGGTATGAAATCATTTTCTTCCTCTGTCACTTCTGTTTTTATCATCAATCT	60
Oy	61	CAGAGCTGACCCACGAATCTTATCAGAGAGCTGTTAAGAACTGGAAAGCTTGGAAAT	120
Db	61	CAGAGCTGACCCACGAATCTTATCAGAGAGCTGTTAAGAACTGGAAAGCTTGGAAAT	120
Oy	121	GCGGTCCTTACCGGATACAGCCTGTTCAGAAAGAGAGATGTCGTGAATGAAGGACG	180
Db	121	GCGGTCCTTACCGGATACAGCCTGTTCAGAAAGAGAGATGTCGTGAATGAAGGACG	180
Oy	181	AGTGGCAGCGCCGTTTAAACACACACCCGGTTCTAAAGGGTCTGTGGTACGGGTGC	240
Db	181	AGTGGCAGCGCCGTTTAAACACACACCCGGTTCTAAAGGGTCTGTGGTACGGGTGC	240
Oy	241	TCCGGTGGGTCGTGGGCTCTGGGGGTTCGTCGACCTCCGGGGGAGCGGTGGCATCAGT	300
Db	241	TCCGGTGGGTCGTGGGCTCTGGGGGTTCGTCGACCTCCGGGGGAGCGGTGGCATCAGT	300
Oy	301	GCGTCAGTGGCAAGCGCGGGGTTCCGGGAACAGTCGAAGAACAATCTCTGCACAACT	360
Db	301	GCGTCAGTGGCAAGCGCGGGGTTCCGGGAACAGTCGAAGAACAATCTCTGCACAACT	360
Oy	361	AGCGATTCCGACGCGCAAGTCTACGCCGACCTCAACACCGGATGAGAAATATCTCCTC	420
Db	361	AGCGATTCCGACGCGCAAGTCTACGCCGACCTCAACACCGGATGAGAAATATCTCCTC	420
Oy	421	ACTATCAGGAGCTGAAGTCCACAGTGTTCGACCTCATTATCATATGCTGCACGTG	480
Db	421	ACTATCAGGAGCTGAAGTCCACAGTGTTCGACCTCATTATCATATGCTGCACGTG	480
Oy	481	TGTGATTAACATTCATGCGCTTCAAAATCTGATTGACGTTTACGAAGAGATCAATGACTC	540
Db	481	TGTGATTAACATTCATGCGCTTCAAAATCTGATTGACGTTTACGAAGAGATCAATGACTC	540
Oy	541	CTGATCAAGTGAATTTCTCTCGACTCGACTGCTTAAGGGCCAACTGAATGACGTTGGCC	600
Db	541	CTGATCAAGTGAATTTCTCTCGACTCGACTGCTTAAGGGCCAACTGAATGACGTTGGCC	600
Oy	601	AATGACTATTGTCCAAATTCATTCATTTGAAGATAGAGCCCAAGAGTGGACGTATTG	660
Db	601	AATGACTATTGTCCAAATTCATTCATTTGAAGATAGAGCCCAAGAGTGGACGTATTG	660
Oy	661	AAGAATTGGTCTTCGGATATCGCAAGCCTCTCGACACATCAAGACAAATGTGGGAAG	720
Db	661	AAGAATTGGTCTTCGGATATCGCAAGCCTCTCGACACATCAAGACAAATGTGGGAAG	720
Oy	721	ATGGAAGATTATTTTAAAGGAATAGAAAGACCATGSAACATTAAGAGCTGATGCAA	780
Db	721	ATGGAAGATTATTTTAAAGGAATAGAAAGACCATGSAACATTAAGAGCTGATGCAA	780
Oy	781	GAATCCAAAAAAGACCATAGACAAATAATAGATGCAACCAAGAGAGAAAGATGAAG	840
Db	781	GAATCCAAAAAAGACCATAGACAAATAATAGATGCAACCAAGAGAGAAAGATGAAG	840
Oy	841	TTGTACCAAGGCCCATGACACTGTCCATCTTATACAAACAGCTTGAAGAGCCCATAC	900
Db	841	TTGTACCAAGGCCCATGACACTGTCCATCTTATACAAACAGCTTGAAGAGCCCATAC	900
Oy	901	CTCATCAGCGAATCGGGAAGGCGCATAGACACCCCTCAAGAAATGAAGAAATATCAAA	960
Db	901	CTCATCAGCGAATCGGGAAGGCGCATAGACACCCCTCAAGAAATGAAGAAATATCAAA	960
Oy	961	CTGCTGCACAAGATTATGAATTAAGAAATCTCCGCCAGCCACTGTGGACACCCCT	1020
Db	961	CTGCTGCACAAGATTATGAATTAAGAAATCTCCGCCAGCCACTGTGGACACCCCT	1020
Oy	1021	AACACGCTGCTGGACAAACAAGAAATAGAGAGACAGAAAGAGATCAAGAGATC	1080
Db	1021	AACACGCTGCTGGACAAACAAGAAATAGAGAGACAGAAAGAGATCAAGAGATC	1080

QY	1081	GCMAAACCTTAAGTTCAACATAATTCCTCTTACTGATCCCTTGAGCTGGAGTAC	1140
Db	1081	GCMAAACCTTAAGTTCAACATAATTCCTCTTACTGATCCCTTGAGCTGGAGTAC	1140
QY	1141	TACTTGAGAGAGAAATATAGAAATATATAGACATCTCCGCAAGTGGAGACAAGAAATCA	1200
Db	1141	TACTTGAGAGAGAAATATAGAAATATATAGACATCTCCGCAAGTGGAGACAAGAAATCA	1200
QY	1201	ACCGAACCTTAATGAATATCCCAATGGTGTGACGTACCCCTCTGTATTAACGATATCAAC	1260
Db	1201	ACCGAACCTTAATGAATATCCCAATGGTGTGACGTACCCCTCTGTATTAACGATATCAAC	1260
QY	1261	AACGCTCTACAGAGACTCAATATAGCTTCGGTGACTTGATTAACCCCTTGATTTACGAAA	1320
Db	1261	AACGCTCTACAGAGACTCAATATAGCTTCGGTGACTTGATTAACCCCTTGATTTACGAAA	1320
QY	1321	GAACCCCTCAAAATATCTACACACAAATGAGAAAGAAAGTTTATACGAAATCTCAAG	1380
Db	1321	GAACCCCTCAAAATATCTACACACAAATGAGAAAGAAAGTTTATACGAAATCTCAAG	1380
QY	1381	GAGAGATCAAAATTTGAGAGAGAAATTTGAGATGTGACAAGAAAAGTTAGAGAACCCG	1440
Db	1381	GAGAGATCAAAATTTGAGAGAGAAATTTGAGATGTGACAAGAAAAGTTAGAGAACCCG	1440
QY	1441	ACCAAAAGCTTAACGATATCTATAAGAGTATGAAAAGCTGTGACAGAGATCTATGAT	1500
Db	1441	ACCAAAAGCTTAACGATATCTATAAGAGTATGAAAAGCTGTGACAGAGATCTATGAT	1500
QY	1501	TCCAAATTCACAAATTAACATGACCTGTGACCAACTTCGAGAAAATGATGGGAAAACGGTAC	1560
Db	1501	TCCAAATTCACAAATTAACATGACCTGTGACCAACTTCGAGAAAATGATGGGAAAACGGTAC	1560
QY	1561	TCTTCAAAAGTGGAGAAACTGACACACATTAATACCTTGCATCCTATGAGAAATCTTAAG	1620
Db	1561	TCTTCAAAAGTGGAGAAACTGACACACATTAATACCTTGCATCCTATGAGAAATCTTAAG	1620
QY	1621	CATAATCTTGAGAGACTCACCAAGCTCTTAAGTATATGAGAGACTATTTCTGCGGAAAC	1680
Db	1621	CATAATCTTGAGAGACTCACCAAGCTCTTAAGTATATGAGAGACTATTTCTGCGGAAAC	1680
QY	1681	ATTGTTTGGAGAAAGAACTAAAGTATTAACAAGATCTCATTAAGTAAAGATGAAAACGAG	1740
Db	1681	ATTGTTTGGAGAAAGAACTAAAGTATTAACAAGATCTCATTAAGTAAAGATGAAAACGAG	1740
QY	1741	ATCGAGAGCTTGTGAGAACACTTAAGAAAGATGAGAAACAGTTGTTTGAGAAAGAAATTT	1800
Db	1741	ATCGAGAGCTTGTGAGAACACTTAAGAAAGATGAGAAACAGTTGTTTGAGAAAGAAATTT	1800
QY	1801	ACAAAAGACGAAATTAACCCAGATGAGAGATCTGTGAGGTCCTCCGATATTTGTTAAAGTC	1860
Db	1801	ACAAAAGACGAAATTAACCCAGATGAGAGATCTGTGAGGTCCTCCGATATTTGTTAAAGTC	1860
QY	1861	CAAGTGCAGAGAGTCTCTCTATGAAACAAGATGATGTAACCTCAAGAACAATCAACTCTTT	1920
Db	1861	CAAGTGCAGAGAGTCTCTCTATGAAACAAGATGATGTAACCTCAAGAACAATCAACTCTTT	1920
QY	1921	CTGAGAAAGCTGTGATTAACATTAATATACATGTGCGCAATAGTTTATTAACAGAGAGAT	1980
Db	1921	CTGAGAAAGCTGTGATTAACATTAATATACATGTGCGCAATAGTTTATTAACAGAGAGAT	1980
QY	1981	AAGCAGGAACCATTAACCTCATCTGTACTCAAGAAAAGATAGACAAAAGTAAAGTGTTC	2040
Db	1981	AAGCAGGAACCATTAACCTCATCTGTACTCAAGAAAAGATAGACAAAAGTAAAGTGTTC	2040
QY	2041	ATGCCCAAAAGTGGAGAGCTGTATCAACGAAGGAAGAAACATTAAAACTGAAGGACAG	2100
Db	2041	ATGCCCAAAAGTGGAGAGCTGTATCAACGAAGGAAGAAACATTAAAACTGAAGGACAG	2100
QY	2101	TCAGATTAATCTCGAGCTTTCACAGAAAGAGATTAACCGGACAGGCTTACCAACAAGGCC	2160
Db	2101	TCAGATTAATCTCGAGCTTTCACAGAAAGAGATTAACCGGACAGGCTTACCAACAAGGCC	2160
QY	2161	GGACAAACAGGCGGTTACGCTTCGAAAGCGATAGCTGTGCAAGCTCAAGCAACGAAGACGAG	2220

|||||
Db 2161 GGACAAACAGCGGTTACACTCTCGAAGGCGATAGGCTCAAGCTCAACACAAAGAGCAG 2220
QY 2221 AAGCAGGACGACCTCCACTGCGAGTGGCCGCTCCAGAGGCTTAAGCTCAAGGCTTACA 2280
Db 2221 AAGCAGGACGACCTCCACTGCGAGTGGCCGCTCCAGAGGCTTAAGCTCAAGGCTTACA 2280
QY 2281 CCACACAGCTCTGTGAATTAACAAGACCGGAATGTGAGCAAACTGACTACCTTGAGAG 2340
Db 2281 CCACACAGCTCTGTGAATTAACAAGACCGGAATGTGAGCAAACTGACTACCTTGAGAG 2340
QY 2341 CTCATGAGTTCCTGAAATACATCTCAATCTGCGACAAATATATCTCTCTCAACG 2400
Db 2341 CTCATGAGTTCCTGAAATACATCTCAATCTGCGACAAATATATCTCTCTCAACG 2400
QY 2401 ACTATGAAAGAGAAATCTTAAACAGTACAAAGATTAACAGAGAGAGAGTAAAGT 2460
Db 2401 ACTATGAAAGAGAAATCTTAAACAGTACAAAGATTAACAGAGAGAGAGTAAAGT 2460
QY 2461 TCCCTCTGTGATCACTGAGACCTGCTGTCAATATCCAGAAACAACTCCGTTATGAT 2520
Db 2461 TCCCTCTGTGATCACTGAGACCTGCTGTCAATATCCAGAAACAACTCCGTTATGAT 2520
QY 2521 TCTATGTTGATAGCTTCAACAATTTCTCTCTCAACTGTTATGAGATATATGAGAG 2580
Db 2521 TCTATGTTGATAGCTTCAACAATTTCTCTCTCAACTGTTATGAGATATATGAGAG 2580
QY 2581 GAGATGCTGCACTGATATTAACATCAAAAGACAAAGACAAAGATTAAGAACTCTGAG 2640
Db 2581 GAGATGCTGCACTGATATTAACATCAAAAGACAAAGACAAAGATTAAGAACTCTGAG 2640
QY 2641 GAAGCTAAGAGGTCCTCACTCTGTGTTAAACTCTCTCTTCCAGCTCCATGCAACCACTG 2700
Db 2641 GAAGCTAAGAGGTCCTCACTCTGTGTTAAACTCTCTCTTCCAGCTCCATGCAACCACTG 2700
QY 2701 TCTCTACACCTCAAGACAAAGCCGAGTGAAGGCTTAAGAGACACCTCTCACTGAGC 2760
Db 2701 TCTCTACACCTCAAGACAAAGCCGAGTGAAGGCTTAAGAGACACCTCTCACTGAGC 2760
QY 2761 AACCTTAATTAACCTGAACTGTTGAGAACATCTCTCTCTGCGCAAGATTAAGAC 2820
Db 2761 AACCTTAATTAACCTGAACTGTTGAGAACATCTCTCTCTGCGCAAGATTAAGAC 2820
QY 2821 ATCTACCAAGAACTTATGAGACAAATGCTCGAAGACTTCTACGAGAGATATCTGAAA 2880
Db 2821 ATCTACCAAGAACTTATGAGACAAATGCTCGAAGACTTCTACGAGAGATATCTGAAA 2880
QY 2881 GACAGGACACATTTCTATACGAGAGCTTCACTAACTTGTGAATCTTAAAGCCGATGAT 2940
Db 2881 GACAGGACACATTTCTATACGAGAGCTTCACTAACTTGTGAATCTTAAAGCCGATGAT 2940
QY 2941 ATCAACTCTTTAAGCATATCTTAAAGCTAAGAGCTGGAAGAGACATCAATAAGCTG 3000
Db 2941 ATCAACTCTTTAAGCATATCTTAAAGCTAAGAGCTGGAAGAGACATCAATAAGCTG 3000
QY 3001 AAGAAAGACCTGCACTGAGCTTCCAGCTGTACACAACTGAACTGAACTGGAAGAGAA 3060
Db 3001 AAGAAAGACCTGCACTGAGCTTCCAGCTGTACACAACTGAACTGAACTGGAAGAGAA 3060
QY 3061 CTCCTGCAAGAAAGAGACAGTGGCAAGTATTAAGATCAGATCAAGAGTGTGACTG 3120
Db 3061 CTCCTGCAAGAAAGAGACAGTGGCAAGTATTAAGATCAGATCAAGAGTGTGACTG 3120
QY 3121 CTCGAAGAGACAGCTGGAAGAACTCACTCACTAAGCAATCCGAACAGCTGACGAG 3180
Db 3121 CTCGAAGAGACAGCTGGAAGAACTCACTCACTAAGCAATCCGAACAGCTGACGAG 3180
QY 3181 AACTTCTCAGTGTCTTCAACAAGAAAGAGAGAGCCGAGATGCGGAGACAGAACT 3240
Db 3181 AACTTCTCAGTGTCTTCAACAAGAAAGAGAGAGCCGAGATGCGGAGACAGAACT 3240
QY 3241 CTGGAGAAACCAAGTCTTCTCAAACTCAAAAGGCTGTCAAGTATATATGAGC 3300
|||||

Db 3241 CTGGAGAAACCAAGATTTCTTCAAACTACAAAGGCTGTCAAGTATATATATGAGC 3300
QY 3301 GAGTCTTCTCTGTGAAGCTCTCTCGAGAGAGATCCAGACCGAGATTAATCACTGCGC 3360
Db 3301 GAGTCTTCTCTGTGAAGCTCTCTCGAGAGAGATCCAGACCGAGATTAATCACTGCGC 3360
QY 3361 AGCTTGAGAACTTCAAGGCTCTGTCTAAGCTGGAAGGCAAGCTGAAGACAACTGAAAC 3420
Db 3361 AGCTTGAGAACTTCAAGGCTCTGTCTAAGCTGGAAGGCAAGCTGAAGACAACTGAAAC 3420
QY 3421 CTGGAGAAAGAGAGCTCAAGTACCTCTCTAGCGAGCTGATACCTGATCCGAGCTC 3480
Db 3421 CTGGAGAAAGAGAGCTCAAGTACCTCTCTAGCGAGCTGATACCTGATCCGAGCTC 3480
QY 3481 AAGAGAGTATTAAAGAAAGAACTACACCGCAATAGCCCAAGAGATTAATACAGAC 3540
Db 3481 AAGAGAGTATTAAAGAAAGAACTACACCGCAATAGCCCAAGAGATTAATACAGAC 3540
QY 3541 GTGAATTAACGCACTGGAATCTTACAAAGAGTCTGCTGGAAGAAACAGATGTCGCACT 3600
Db 3541 GTGAATTAACGCACTGGAATCTTACAAAGAGTCTGCTGGAAGAAACAGATGTCGCACT 3600
QY 3601 GTGGTGTGGAATCTGGGCTCCAGACACAGAGAGCACTGCACTCACTTAAGAGCTGATCT 3660
Db 3601 GTGGTGTGGAATCTGGGCTCCAGACACAGAGAGCACTGCACTCACTTAAGAGCTGATCT 3660
QY 3661 ACTCATGTGAGCGGAGCGAGTCCAAATACATTAACATCACTCAGAACCTGACAGTGAAGTTC 3720
Db 3661 ACTCATGTGAGCGGAGCGAGTCCAAATACATTAACATCACTCAGAACCTGACAGTGAAGTTC 3720
QY 3721 GATGAGCTATATTTGCTGCTATCTTCCGCGAGAGAGGAGAGACTACATGACCTGCGC 3780
Db 3721 GATGAGCTATATTTGCTGCTATCTTCCGCGAGAGAGGAGAGACTACATGACCTGCGC 3780
QY 3781 CAGGTGTGAGAGGCTGCACTCTCTCCGATGATTAATCACTTCTGTCGCAAAATC 3840
Db 3781 CAGGTGTGAGAGGCTGCACTCTCTCCGATGATTAATCACTTCTGTCGCAAAATC 3840
QY 3841 GAGAAAGAAATAGCAAGTGTCTATCTGAAACCTTGAGAGGCGTATAGTCTCTCAAG 3900
Db 3841 GAGAAAGAAATAGCAAGTGTCTATCTGAAACCTTGAGAGGCGTATAGTCTCTCAAG 3900
QY 3901 AAAACAGCTGGAATTAACCTGATGACCTTCAATGCAAGTGAAGGACATTCGAAACAGC 3960
Db 3901 AAAACAGCTGGAATTAACCTGATGACCTTCAATGCAAGTGAAGGACATTCGAAACAGC 3960
QY 3961 CGCTTAATTAAGAGAAATTTCAAGAGGCTTGGAGAGGCACTTGTCCCTATAAA 4020
Db 3961 CGCTTAATTAAGAGAAATTTCAAGAGGCTTGGAGAGGCACTTGTCCCTATAAA 4020
QY 4021 GACCTGACCTCCTTAAGTACCTGTTGCAAGAGCCATACAGTCTCTCAATAAGAGAG 4080
Db 4021 GACCTGACCTCCTTAAGTACCTGTTGCAAGAGCCATACAGTCTCTCAATAAGAGAG 4080
QY 4081 AAGGATTAATTTCTGTCTGTTACAACTATATCAAGAGGCTGACAGCCGATCAAT 4140
Db 4081 AAGGATTAATTTCTGTCTGTTACAACTATATCAAGAGGCTGACAGCCGATCAAT 4140
QY 4141 TTGCGTAATGATGTGCTGGGATTTACAAAGATCTGAGCGAAATATCAAGTCTGACTT 4200
Db 4141 TTGCGTAATGATGTGCTGGGATTTACAAAGATCTGAGCGAAATATCAAGTCTGACTT 4200
QY 4201 GACTCTATTAAAAAGTATATCAACGATTAAGCAAGGCGAGATGAATATCTGCCCTTC 4260
Db 4201 GACTCTATTAAAAAGTATATCAACGATTAAGCAAGGCGAGATGAATATCTGCCCTTC 4260
QY 4261 CTGAATTAACATGGAACCTGTACAAAGAGAGAGAGAGCAAAATATCACTCTTCTGTAAT 4320
Db 4261 CTGAATTAACATGGAACCTGTACAAAGAGAGAGAGAGCAAAATATCACTCTTCTGTAAT 4320
QY 4321 CACCTGAGAGCCCAAGGCTCTCAACTATCTTACGAGAGAGCAATGTGAAGTTAAATC 4380
Db 4321 CACCTGAGAGCCCAAGGCTCTCAACTATCTTACGAGAGAGCAATGTGAAGTTAAATC 4380

```
QY 4381 AAGGAGCTGAACCTACCTCAAAAACATCCAGACAGCTGGAGATTGCAAGAAAATAC 4440
Db 4381 AAGGAGCTGAACCTACCTCAAAAACATCCAGACAGCTGGAGATTGCAAGAAAATAC 4440
QY 4441 AATTTCGTGGAAATTCAGACCTGTCTACCGATTATACCAACCAATCTCTGACCAAG 4500
Db 4441 AATTTCGTGGAAATTCAGACCTGTCTACCGATTATACCAACCAATCTCTGACCAAG 4500
QY 4501 TTTCTGTCCACTGGCAGTGGTGTTCGAAAACCTGCCAAAACAGTGTGACCAATCTGCTC 4560
Db 4501 TTTCTGTCCACTGGCAGTGGTGTTCGAAAACCTGCCAAAACAGTGTGACCAATCTGCTC 4560
QY 4561 GACGGCAACCTGCAAGGGCATCTGAACTCCCGACGACCAATGGCTGAAGAAACAGTGC 4620
Db 4561 GACGGCAACCTGCAAGGGCATCTGAACTCCCGACGACCAATGGCTGAAGAAACAGTGC 4620
QY 4621 CCCGAGAAATAGCGGCTGTTCAGGCATCTGGACGACGCGAAGAGTGAAGTGTCTCCTG 4680
Db 4621 CCCGAGAAATAGCGGCTGTTCAGGCATCTGGACGACGCGAAGAGTGAAGTGTCTCCTG 4680
QY 4681 AACTACAAACAAGAGAGATAGTGCCTGGAGAACCCAAACCTGCAATGAAAAC 4740
Db 4681 AACTACAAACAAGAGAGATAGTGCCTGGAGAACCCAAACCTGCAATGAAAAC 4740
QY 4741 AATGGCGGGTGTGACGGCGATGCTAAATGACACCGAGAAAGACAGGCGCTTAACGGAAG 4800
Db 4741 AATGGCGGGTGTGACGGCGATGCTAAATGACACCGAGAAAGACAGGCGCTTAACGGAAG 4800
QY 4801 AAAATCAACATCGAGTGTACTAAGCCGACACTCTATCCACTCTTCGACGGGATTTTTCG 4860
Db 4801 AAAATCAACATCGAGTGTACTAAGCCGACACTCTATCCACTCTTCGACGGGATTTTTCG 4860
QY 4861 TTCAGCTTAAATTTCTCTGGGATCTCCTTCCTGCTGATTCCTATCTGATCTGTACAGC 4920
Db 4861 TTCAGCTTAAATTTCTCTGGGATCTCCTTCCTGCTGATTCCTATCTGATCTGTACAGC 4920
QY 4921 TTCATCTAATAGATGATGG 4940
Db 4921 TTCATCTAATAGATGATGG 4940

RESULT 2
CVEI31294 4940 bp DNA Linear SYN 24-FEB-1999
LOCUS Cloning vector pBSK*-MSP-1s/FCB.
ACCESSION AJ131294
VERSION AJ131294.1 GI:4455908
KEYWORDS Cloning vector; major surface protein-1; msp-1 gene.
SOURCE Cloning vector pBSK*-MSP-1s/FCB.
ORGANISM artificial sequences; vectors.
REFERENCE 1 (bases 1 to 4940)
AUTHORS Bujard,H.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1998) Bujard H., ZMBH - Zentrum fuer Molekulare
Biologie, Universitaet Heidelberg, Im Neuenheimer Feld 282, 69120
Heidelberg, BW, GERMANY
REFERENCE 2 (bases 1 to 4940)
AUTHORS Pan,W., Kayot,E., Tolle,R., Frank,R., Mosbach,R., Turbachova,I. and
Bujard,H.
TITLE Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned
4917 bp polynucleotide enables synthesis and isolation of
full-length protein from Escherichia coli and mammalian cells
JOURNAL Nucleic Acids Res. 27 (4), 1094-1103 (1999)
MEDLINE 99128299
PUBMED 99277744
FEATURES
source location/Qualifiers
1..4940
/organism="Cloning vector pBSK*-MSP-1s/FCB"
/db_xref="taxon:85834"
1..4940
source
```

```
/organism="Plasmodium falciparum"
/strain="FCB-1"
/db_xref="taxon:5833"
10..4929
/gene="msp-1"
10..4929
/gene="msp-1"
10..4929
/gene="msp-1"
/function="surface protein"
/codon_start=1
/evidence="experimental"
/product="major surface protein-1"
/protein_id="CA836903.1"
/db_xref="GI:4455909"
/translacion="MKTIPLCSPLFLIMQCTHESVOELVKKLEALBDAYLTVGS
LFOKEKAVLNBGTSIGTAVTTSTPSGSKSVASGSGGVSASGVSASGVSAS
GGSGNSRRTPNSDSQSDAKSYADLHNRVNNYLTKLEKLPOLFDLTNLTCDN
IHGFKYLIDGEEELNLYKLNFEDLIRAKLNDVNCANDYCOIPEFNKIRANLEDVLR
KLVGKRPPLDNIKDNVGMEDYIKKKKTLEINLEIESKTIIDKNKATKEEEK
KLVOAYDLSTYNKLEAHNLISLEKRIIDTLKKNENIKELDKINENKPPANG
NTPWTLDDKKKIIEHEKEIKELAKTIKFNIDSLFMDPLEVLYLREKNNIDISARV
ETKESTEPNEVPNCVTPPLSYNDINNALNELNSFGDILNPDTIKESKNIYDNEK
KFLNEIKETIKETKLESOKSYEDSKSLNDITTEKLELNTIDSKNNNIDLTN
FEKMMGRYSYKEVEKLHNHTFASVENSKNLEKLTALVMEKYSLRNIYVEKELY
YKNLSIKLENIETLVENIRKDEOLEFEKTIKDEKNPDEKILESDIVKVOYKVL
MKNIDELKTIOLILKVELKNIHVPSYSEONKOEYLYILVLEKLEIDKLVEMPKYE
SLINEEKNITKTESONSSEPSSTEGEITGQATTPRGOAASLAEGSVOAOBOKA
QPPYPPVPEKAKOYTPPAPVNNKTENSTYDLEKLEPLNTSYICHTIYLSHT
MNEKILKYOYITKEESKLSQDPLDLPFNONNI PWYSMFDSLNSLSQLEIYTE
KEWYCNLYKLDNDKIKNLEAKRYESTVSTSSSMQPSLTPDOKPVSADNTS
HSTNNNSLKLFEINLSLGNKNIYQELIQKSSSENFYERILKSDPFYSEFPNFK
SKADDISLNDSESKRKLIEDINKLTLOLSPDYLYKRYKLELFDKKTVGKYK
OIKRTLLKLELESKRKLSLNNPKHVLQNFVSFPMKKKEALIAENTLEMPKILKRYK
KGLKYKYNNGESSPLKTLSESIOTREDYASILENKVYSLEKRLKNDLNLEKRLSTL
SSGLHLILAEKEVTKNNKTGNPSNENVDNNALSIKFLBEGDVTATVSESS
DTLESOQPKRPASTHVAESNTITTSQNVDEVDVITVPIGSESDYDLDGVVGS
EAVPVSVIDMLSKILENEVLYLKLPLAGVYRSLLKOLENNVMFNNAVVDILNSRN
KRENFKNVLESDDLPRKDLTSSNVVDPYFLKLEKREKFLSSSYNIKDSIDTDFN
ANDVLYGKIILSEKYSKSDLSIKKYLINDKQENKRYPLFNIEITLYKTVNDKIDLTV
IHLEAKVLYTYEKSNEVYKIKELNYLKTIODKADKKNNFPGIADLTSDVNNHL
LTKRLSTGWMPENIAKTIVLSNLDGNLQGHNLISQHCYKQCCQNGSCRRHDERE
CKCLINTKQEGDKCVENPNPFCNNNGGCDADACTEEDSGSNKKTTCCTCPDSTP
LFDDIFCSSNFILFIPLILMLILYSFI"
10..69
/gene="msp-1"
sig_peptide
BASE COUNT 1669 a 1156 c 1060 g 1055 t
ORIGIN
Query Match 100.0%; Score 4938.4; DB 12; Length 4940;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CGCAGCGCTAAGAAATATATTTCTCTCTGCTGATCTTCGTTTATATCATCAATCT 60
Db 1 CGCAGCGCTAAGAAATATATTTCTCTCTGCTGATCTTCGTTTATATCATCAATCT 60
QY 61 CAGTCCGTGACCCACGAATCTATCAGAGCTGTTAAGAACTGGAAGAT 120
Db 61 CAGTCCGTGACCCACGAATCTATCAGAGCTGTTAAGAACTGGAAGAT 120
QY 121 GCCGCTCTTACCGGATACAGCTGTTCCAGAAAGGAAAGATGAGGAGACG 180
Db 121 GCCGCTCTTACCGGATACAGCTGTTCCAGAAAGGAAAGATGAGGAGACG 180
QY 181 AGTGGCAGCGCGCTTACACAGACACACCGCGTTCTAAAGGCTGTGAGCGGTGC 240
Db 181 AGTGGCAGCGCGCTTACACAGACACACCGCGTTCTAAAGGCTGTGAGCGGTGC 240
QY 241 TCCGATGAGTCTGTGGCTGTGGGGGTTCCGCTCCCGGCGACGAGTGGCATCAGGT 300
Db 241 TCCGATGAGTCTGTGGCTGTGGGGGTTCCGCTCCCGGCGACGAGTGGCATCAGGT 300
QY 301 GGCTCAGTGGCAGCGGCGGTTCCGGGAAAGTGAAGAACCAATCCATCTGACACTCT 360
Db 301 GGCTCAGTGGCAGCGGCGGTTCCGGGAAAGTGAAGAACCAATCCATCTGACACTCT 360
```

```

Db 301 GGCCTAGTGGCAAGCGCGGTTCCGGGAACGTGAGAACCAATCCATCTGACAACTCT 360
QY 361 ACCGATTCGAGCCCAAGTCTACGCCGACCTCAACGACCAGTGGAGAAATATCTCTC 420
Db 361 ACCGATTCGAGCCCAAGTCTACGCCGACCTCAACGACCAGTGGAGAAATATCTCTC 420
QY 421 ACTATCAAGAGAGCTGAGAGTACCCACAGTTGTTGACCTCAGTAATCATATGCTGACACTG 480
Db 421 ACTATCAAGAGAGCTGAGAGTACCCACAGTTGTTGACCTCAGTAATCATATGCTGACACTG 480
QY 481 TGTGATTAACATTCATGAGCTTCAAAATATCGATTTGAGGTTAGGAAAGATCAATGAACTC 540
Db 481 TGTGATTAACATTCATGAGCTTCAAAATATCGATTTGAGGTTAGGAAAGATCAATGAACTC 540
QY 541 CTGTCAAGATTTGAATTTCTACTGCTGCTGAAGGCCAAAGTGAATGACGTTTCCGCC 600
Db 541 CTGTCAAGATTTGAATTTCTACTTCTGACTTCTGATGAGGCCAAAGTGAATGACGTTTCCGCC 600
QY 601 AATGACTATTTGCAAAATTTCCATTTCAATTTGAAGATCAGAGCCAAAGGTTGACGTTTGG 660
Db 601 AATGACTATTTGCAAAATTTCCATTTCAATTTGAAGATCAGAGCCAAAGGTTGACGTTTGG 660
QY 661 AAGAAGTTGCTCTGGATATCGCAAGCCTCTGACACATCAAGACATGCTGGCAAG 720
Db 661 AAGAAGTTGCTCTGGATATCGCAAGCCTCTGACACATCAAGACATGCTGGCAAG 720
QY 721 ATGGAAGATTTATTTAAAAAGATTAAGAAGACCATGAGACATTTAGACGCTGATCGAA 780
Db 721 ATGGAAGATTTATTTAAAAAGATTAAGAAGACCATGAGACATTTAGACGCTGATCGAA 780
QY 781 GAATCCAAAAGAACCATATAGCAAAAAATAGAAATGCAACCAAGAGAGAAAAAGAAAG 840
Db 781 GAATCCAAAAGAACCATATAGCAAAAAATAGAAATGCAACCAAGAGAGAAAAAGAAAG 840
QY 841 TTGTACACAGAGCCAGTACGACAGCTGTCTCATCTATTAACAAAGCTTTAAGAAAGCCCTAAC 900
Db 841 TTGTACACAGAGCCAGTACGACAGCTGTCTCATCTATTAACAAAGCTTTAAGAAAGCCCTAAC 900
QY 901 CTCATCAGAGCTACTGAGAGAGCGCATAGACACCTCAAGACAGATGAATATCAAGAA 960
Db 901 CTCATCAGAGCTACTGAGAGAGCGCATAGACACCTCAAGACAGATGAATATCAAGAA 960
QY 961 CTGCTCGACAGATTTAATGAATTAAGAAATCCTCGCAGCCAACTCTGGGAAACACCCCT 1020
Db 961 CTGCTCGACAGATTTAATGAATTAAGAAATCCTCGCAGCCAACTCTGGGAAACACCCCT 1020
QY 1021 AACAGGCTCTGACAAAGAAAGAAAGATAGAGAGCAGAGAGAAAGATCAAAAGATC 1080
Db 1021 AACAGGCTCTGACAAAGAAAGAAAGATAGAGAGCAGAGAGAAAGATCAAAAGATC 1080
QY 1081 GCCAAAGCATTAAAGTTCAACATAGATCTCTCTTACTGATCCCTTGAGCTGGAGTAC 1140
Db 1081 GCCAAAGCATTAAAGTTCAACATAGATCTCTCTTACTGATCCCTTGAGCTGGAGTAC 1140
QY 1141 TACTTGAGAGAGAAAGATTAAGAAATATAGACATCTCCGCCAAAGTCCAGACAAAGGATCA 1200
Db 1141 TACTTGAGAGAGAAAGATTAAGAAATATAGACATCTCCGCCAAAGTCCAGACAAAGGATCA 1200
QY 1201 ACCGAAGCTTAAGAAATCCCAATGAGTGTGACGTACCTCTGCTTATTAACGATATCAAC 1260
Db 1201 ACCGAAGCTTAAGAAATCCCAATGAGTGTGACGTACCTCTGCTTATTAACGATATCAAC 1260
QY 1261 AACGCTCTAAGAGAGCTCAATAGCTCGGTGACTGATTAACCCCTTGATTAAGAA 1320
Db 1261 AACGCTCTAAGAGAGCTCAATAGCTCGGTGACTGATTAACCCCTTGATTAAGAA 1320
QY 1321 GAACCTCTAAGAGATTTCTACACAGACAAATGAGAGAAAGAGTTTATCAACGAAATCAAG 1380
Db 1321 GAACCTCTAAGAGATTTCTACACAGACAAATGAGAGAAAGAGTTTATCAACGAAATCAAG 1380
QY 1381 GAGAGATCAAAATTTGAGAGAAAGAAATTTGAGAGTGAACAAAGATTTACGAAGCCG 1440
Db 1381 GAGAGATCAAAATTTGAGAGAAAGAAATTTGAGAGTGAACAAAGATTTACGAAGCCG 1440
QY 1441 AGCAAAAGTCTAAAGATATCACTTAAGAGATTAAGAAAGCTCTGTAAGAGATCTATGAT 1500
Db 1441 AGCAAAAGTCTAAAGATATCACTTAAGAGATTAAGAAAGCTCTGTAAGAGATCTATGAT 1500
QY 1501 TCCAAATTTACAAATTAACATTCGACCTGACCAACTTGAAGAAATGATGGGAAACGCTAC 1560
Db 1501 TCCAAATTTACAAATTAACATTCGACCTGACCAACTTGAAGAAATGATGGGAAACGCTAC 1560
QY 1561 TCTTACAAAGTGGAGAAAGTGAACACACCATTAATACCTTTGCATTCCTATGAGAAATTTAG 1620
Db 1561 TCTTACAAAGTGGAGAAAGTGAACACACCATTAATACCTTTGCATTCCTATGAGAAATTTAG 1620
QY 1621 CATTAATCTTGAGAGCTCAACAAAGCTCTTAAGTAAATGAGAGACTATTTCTGCGGAAC 1680
Db 1621 CATTAATCTTGAGAGCTCAACAAAGCTCTTAAGTAAATGAGAGACTATTTCTGCGGAAC 1680
QY 1681 ATTGTTGTGAGAGAAAGACTTAAGATTTTACAAAGATCTATAAGTATGCAAAAAGAG 1740
Db 1681 ATTGTTGTGAGAGAAAGACTTAAGATTTTACAAAGATCTATAAGTATGCAAAAAGAG 1740
QY 1741 ATCGAGAGCTGTTGAGAAACATTAAGAGATTAAGAAAGATGAAAGAGTGTGTAAGAAAGAT 1800
Db 1741 ATCGAGAGCTGTTGAGAAACATTAAGAGATTAAGAAAGATGAAAGAGTGTGTAAGAAAGAT 1800
QY 1801 ACAAAAGAGAAATTAACACAGATGAGAGATCTGAGAGCTCCGATATTTGTTAAAGTC 1860
Db 1801 ACAAAAGAGAAATTAACACAGATGAGAGATCTGAGAGATCTGAGAGCTCCGATATTTGTTAAAGTC 1860
QY 1861 CAAGTCAGAGAGTGTCTCTCATGAACAAGATTTGAATCTCAAGAGACTCAACATCATT 1920
Db 1861 CAAGTCAGAGAGTGTCTCTCATGAACAAGATTTGAATCTCAAGAGACTCAACATCATT 1920
QY 1921 CTGAAGAGAGTGGAGTTAAACATTAATATACATGTCGCGAATAGTATTAAGCAGAGAAAT 1980
Db 1921 CTGAAGAGAGTGGAGTTAAACATTAATATACATGTCGCGAATAGTATTAAGCAGAGAAAT 1980
QY 1981 AAGCAGAGAACATTAATACCTCTGATCAAGAAAGAGATAGACAACTGAAAGTGTTC 2040
Db 1981 AAGCAGAGAACATTAATACCTCTGATCAAGAAAGAGATAGACAACTGAAAGTGTTC 2040
QY 2041 ATGCCCAAAGTGAAGAGCTGATCAACGAGAGAGAGAAAGAACTTAAGAAAGTGAAGAGAG 2100
Db 2041 ATGCCCAAAGTGAAGAGCTGATCAACGAGAGAGAGAAAGAACTTAAGAAAGTGAAGAGAG 2100
QY 2101 TCAGATTAATCCGAGGCTTCCACAGAAAGAGAGATTAACCGGACAGGCTACCAACAGGCC 2160
Db 2101 TCAGATTAATCCGAGGCTTCCACAGAAAGAGAGATTAACCGGACAGGCTACCAACAGGCC 2160
QY 2161 GGACAAAGAGCGGTTACAGCTCTGAAAGCGATAGCGTCAAGCTCAAGCACAAGAGCAG 2220
Db 2161 GGACAAAGAGCGGTTACAGCTCTGAAAGCGATAGCGTCAAGCTCAAGCACAAGAGCAG 2220
QY 2221 AAGCAGGACAGGCTCTCAATGCGCAGTGGCCGTTCCAGAGGCTTAAGACTCAAGTGGCTTAA 2280
Db 2221 AAGCAGGACAGGCTCTCAATGCGCAGTGGCCGTTCCAGAGGCTTAAGACTCAAGTGGCTTAA 2280
QY 2281 CCAACGAGTCTGTGATTAACAAAGACCGAGAAATGTAGCAAACTGAGTACTTGAAG 2340
Db 2281 CCAACGAGTCTGTGATTAACAAAGACCGAGAAATGTAGCAAACTGAGTACTTGAAG 2340
QY 2341 CTCTATGAGTTCCTGAATACATCTACATCTGCGACAAATATATCTGCTCTCTCACAGC 2400
Db 2341 CTCTATGAGTTCCTGAATACATCTACATCTGCGACAAATATATCTGCTCTCTCACAGC 2400
QY 2401 ACTATGAAGAGAGATTTCTTAACAGTACAGATTAACCAAGAGAGAGAGAGTAAAGT 2460
Db 2401 ACTATGAAGAGAGATTTCTTAACAGTACAGATTAACCAAGAGAGAGAGTAAAGT 2460
QY 2461 TCTCTTGTGATCCACTGAGACCGTGTTCATATTCAGAAACAACTCCCGTTATGTAT 2520
Db 2461 TCTCTTGTGATCCACTGAGACCGTGTTCATATTCAGAAACAACTCCCGTTATGTAT 2520

```

Oy	2521	TCCTGTTCCAGTATGCCCTCAACAACTTCTCTCTCTCAACAGCTGTATGGAGATATATGAGAG	2580
Db	2521	TCCTATGTTCCAGTATGCCCTCAACAACTTCTCTCTCTCAACAGCTGTATGGAGATATATGAGAG	2580
Oy	2581	GAGATGGTCTGCAACCGTATTAACCTCAAGAACAGCAAGCAATTAATAGAACCTTCTGGAG	2640
Db	2581	GAGATGGTCTGCAACCGTATTAACCTCAAGAACAGCAAGCAATTAATAGAACCTTCTGGAG	2640
Oy	2641	GAAAGCTAAGAGAGGTCTCCACCTCTGTTPAAAACCTCTCTCTCCAGCTCCAGTCCAACCACTG	2700
Db	2641	GAAAGCTAAGAGAGGTCTCCACCTCTGTTPAAAACCTCTCTCTCCAGCTCCAGTCCAACCACTG	2700
Oy	2701	TCCTTCACACCTCAAGACCAAGCCCCGAAGAGGCGCTAACAGACAGCACTCTCAGCCAGC	2760
Db	2701	TCCTTCACACCTCAAGACCAAGCCCCGAAGAGGCGCTAACAGACAGCACTCTCAGCCAGC	2760
Oy	2761	AACCTTATATACCTACAGAACTGTTTGAAACATCTCTCTCTGGCAAGATATAGAAC	2820
Db	2761	AACCTTATATACCTACAGAACTGTTTGAAACATCTCTCTCTGGCAAGATATAGAAC	2820
Oy	2821	ATCTACCAAGAACTTATTTGGACAGAAATCGTCGGAACCTTCTACAGAGAGTACTGAAA	2880
Db	2821	ATCTACCAAGAACTTATTTGGACAGAAATCGTCGGAACCTTCTACAGAGAGTACTGAAA	2880
Oy	2881	GACAGCCGACACATTTCTATTAACGAGACCTTCACTAATCTGTGAATCTTAAAGCCGATGAT	2940
Db	2881	GACAGCCGACACATTTCTATTAACGAGACCTTCACTAATCTGTGAATCTTAAAGCCGATGAT	2940
Oy	2941	ATCAACCTCTTAAACGATGAATCTTAAACGTAAGAGCTGGAAGAGCAATCAATTAAGCTG	3000
Db	2941	ATCAACCTCTTAAACGATGAATCTTAAACGTAAGAGCTGGAAGAGCAATCAATTAAGCTG	3000
Oy	3001	AAGAAGACACTGCAACTGACCTTCGACCTGTCAACAACTGACAAACTGAAACTGGAGAGA	3060
Db	3001	AAGAAGACACTGCAACTGACCTTCGACCTGTCAACAACTGACAAACTGAAACTGGAGAGA	3060
Oy	3061	CTCTTCGACAAAGAAAGAGACAGTCGGCAGTATAGATGAGATCAAGAACTTGACTCTG	3120
Db	3061	CTCTTCGACAAAGAAAGAGACAGTCGGCAGTATAGATGAGATCAAGAACTTGACTCTG	3120
Oy	3121	CTCAAGAGAGAGCTTGAAAGCAACCTCACTCACTGATCAATCCGAAACACGTACTGAG	3180
Db	3121	CTCAAGAGAGAGCTTGAAAGCAACCTCACTCACTGATCAATCCGAAACACGTACTGAG	3180
Oy	3181	AACCTTCAGATGTTCTTCAACAAGAAAGAAAGCCGAGATCGCCGAGACAGAGAAACCT	3240
Db	3181	AACCTTCAGATGTTCTTCAACAAGAAAGAAAGCCGAGATCGCCGAGACAGAGAAACCT	3240
Oy	3241	CTGAGAGAAACCAAGATTTCTTCAAAACCTCAAAAGGCTCTGTCAMTATTTATAATAGGC	3300
Db	3241	CTGAGAGAAACCAAGATTTCTTCAAAACCTCAAAAGGCTCTGTCAMTATTTATAATAGGC	3300
Oy	3301	GAGCTTCCTCTCTGAAAGACTCTCTCCGAGGAGAGATCCAGACCGAGAGATCACTACGCC	3360
Db	3301	GAGCTTCCTCTCTGAAAGACTCTCTCCGAGGAGAGATCCAGACCGAGAGATCACTACGCC	3360
Oy	3361	AGCCTCGAGAACTTCAAGGTCCTGTTAAAGCTCGAAGGCAAGCTTAAAGACAACTGAGAC	3420
Db	3361	AGCCTCGAGAACTTCAAGGTCCTGTTAAAGCTCGAAGGCAAGCTTAAAGACAACTGAGAC	3420
Oy	3421	CTGAGAGAAAGAGCTCAGCTACCTCTCTAGCGGACGTGACATCACCTGATCGCGAGATC	3480
Db	3421	CTGAGAGAAAGAGCTCAGCTACCTCTCTAGCGGACGTGACATCACCTGATCGCGAGATC	3480
Oy	3481	AAGGAATCAATTAAGAAACAAGAACTACACCGGCAATATAGCCCAAGGAGAAATAATACAGAC	3540
Db	3481	AAGGAATCAATTAAGAAACAAGAACTACACCGGCAATATAGCCCAAGGAGAAATAATACAGAC	3540
Oy	3541	GTGAAATTAACCACTGGAATTTTACAAGAAAGTTCTGCTCGAAGGAGACAGATGTGCCACT	3600
Db	3541	GTGAAATTAACCACTGGAATTTTACAAGAAAGTTCTGCTCGAAGGAGACAGATGTGCCACT	3600
Oy	3601	GTGGTGTCTGAATCTGGCTCCGACACACTGAGGAGCTCTCAACCTTAAGAAAGCTCTGACTCT	3660

Db	3601	GTGGGTCTGAATCTGGGCTCGACACACTGGAGCGCTCAACCTTAAGAACCTTCGACATCT	3660
Oy	3661	ACTCATGTGCGAGCCGAGTCCAATATCAATTTACCAATCTCAAGAACGTGGAGATGAGTTC	3720
Db	3661	ACTCATGTGCGAGCCGAGTCCAATATCAATTTACCAATCTCAAGAACGTGGAGATGAGTTC	3720
Oy	3721	GATGACGTTCATCTTGTGCTCTCTCTTGGCGAGACGAGAGAGACTACATGACTCGGC	3780
Db	3721	GATGACGTTCATCTTGTGCTCTCTCTTGGCGAGACGAGAGAGACTACATGACTCGGC	3780
Oy	3781	CAGGGTGTCCACGGGTGAGAGCTGTACCTCTTCCTCGATTTGATTAACATTTCTGCCAAATTC	3840
Db	3781	CAGGGTGTCCACGGGTGAGAGCTGTACCTCTTCCTCGATTTGATTAACATTTCTGCCAAATTC	3840
Oy	3841	GAGAACGAATACGAGAGTCTCTATTTGAAACCTCTGGCAGCGTCTATAGGTCTCTCAAG	3900
Db	3841	GAGAACGAATACGAGAGTCTCTATTTGAAACCTCTGGCAGCGTCTATAGGTCTCTCAAG	3900
Oy	3901	AAACAGCTGGAGAAATACGTTGATGACCTTCAATATGCAACGTGAAGACATTTCTGAACGC	3960
Db	3901	AAACAGCTGGAGAAATACGTTGATGACCTTCAATATGCAACGTGAAGACATTTCTGAACGC	3960
Oy	3961	CCCTTTAATTAAGAGAGAAATTTTCAAGAACGCTTGGAGAGCGACTTGTATTTCCCTATAAA	4020
Db	3961	CCCTTTAATTAAGAGAGAAATTTTCAAGAACGCTTGGAGAGCGACTTGTATTTCCCTATAAA	4020
Oy	4021	GACCTGACCTCCTCTACTACGTTTCAAGAGCCCATACAAAGTTCTCTCAATTAAGAGAG	4080
Db	4021	GACCTGACCTCCTCTACTACGTTTCAAGAGCCCATACAAAGTTCTCTCAATTAAGAGAG	4080
Oy	4081	AGGGATTAATTTCTGTCTAGTTTACAACTATATCAAGACCTCCATGACACCGATATCAAT	4140
Db	4081	AGGGATTAATTTCTGTCTAGTTTACAACTATATCAAGACCTCCATGACACCGATATCAAT	4140
Oy	4141	TTTCGCTAATGATGTCTCTGGGATTTACAAAGATCCTGACCGAGAAATTAACAGTCTGACCTT	4200
Db	4141	TTTCGCTAATGATGTCTCTGGGATTTACAAAGATCCTGACCGAGAAATTAACAGTCTGACCTT	4200
Oy	4201	GACCTTATTAATAAATATATATCAACGATTAAGCAAGCGCAGAAATGAATAATATCTGCCCTTC	4260
Db	4201	GACCTTATTAATAAATATATATCAACGATTAAGCAAGCGCAGAAATGAATAATATCTGCCCTTC	4260
Oy	4261	CTGATTAACATCGAAGAACCTCTGACAAAGACAGTGAAGCAACAAATTCGACTCTTGTGTAAT	4320
Db	4261	CTGATTAACATCGAAGAACCTCTGACAAAGACAGTGAAGCAACAAATTCGACTCTTGTGTAAT	4320
Oy	4321	CACCTTGGAGGCCAAGGTCTCTCAACTATCTTACGAAAGACCAATGTGGAAGTTAAAAATC	4380
Db	4321	CACCTTGGAGGCCAAGGTCTCTCAACTATCTTACGAAAGACCAATGTGGAAGTTAAAAATC	4380
Oy	4381	AAGGAGCTGAACCTACTCCAAAACATTCACAGCAAGCTGGAGATTTCAAGAAAAAATTAAC	4440
Db	4381	AAGGAGCTGAACCTACTCCAAAACATTCACAGCAAGCTGGAGATTTCAAGAAAAAATTAAC	4440
Oy	4441	AATTTCTGTGCAATTTGACAGCCTGTCTACCGATTAATAACCAACAAATCTCTGACCAAG	4500
Db	4441	AATTTCTGTGCAATTTGACAGCCTGTCTACCGATTAATAACCAACAAATCTCTGACCAAG	4500
Oy	4501	TTTCTGTCTCACTGCGCATGTGTTGCAAAAACCTCGCCAAACAGTGTGAGCAATCTGCTC	4560
Db	4501	TTTCTGTCTCACTGCGCATGTGTTGCAAAAACCTCGCCAAACAGTGTGAGCAATCTGCTC	4560
Oy	4561	GACGGCAACCTGCAAGGAGATGCTGAACATCTTCCACGACCAACAAATGGTGAACAAACAGTGC	4620
Db	4561	GACGGCAACCTGCAAGGAGATGCTGAACATCTTCCACGACCAACAAATGGTGAACAAACAGTGC	4620
Oy	4621	CCCCGAATTAACGCGCTGTGTTTCAGGCATTTGACGACGCGCAAGAGTGTGMAATGTCTCTCG	4680
Db	4621	CCCCGAATTAACGCGCTGTGTTTCAGGCATTTGACGACGCGCAAGAGTGTGMAATGTCTCTCG	4680
Oy	4681	AACCTCAACAAGAGAGATTAAGTGTGTGAGAACCCAAACCTTACCTGCAATGAAATC	4740

D6	4661	AACGACAAACAAGAGAGATRAAGTGCGCTGGAGAACCACCAACCCTACTGTGCATGAATGAAC	4740
OY	4741	AATGGCGGGTGTGACGCCGATGTCAATATGCACCCGAGAGAACAGCGGCTTTAACGGAAAG	4800
D6	4741	AATGGCGGGTGTGACGCCGATGTCAATATGCACCCGAGAGAACAGCGGCTTTAACGGAAAG	4800
OY	4801	AAAATCACATCGCATGTACTAAGCCGACGTCATTCCACGCTTCGACGGGATTTTTTGGC	4860
D6	4801	AAAATCACATCGCATGTACTAAGCCGACGTCATTCCACGCTTCGACGGGATTTTTTGGC	4860
OY	4861	TCCAGCTCTAAATTTTCCCTGGGATCTCTCTCCCTGCGATCTCATATGTCGATCTCTGACAC	4920
D6	4861	TCCAGCTCTAAATTTTCCCTGGGATCTCTCTCCCTGCGATCTCATATGTCGATCTCTGACAC	4920
OY	4921	TTTCATCTAATAGATCGATGG	4940
D6	4921	TTTCATCTAATAGATCGATGG	4940
RESULT_3			
LOCUS	PRANT195	5917 bp	mRNA linear INV 30-MAR-1995
DEFINITION	Plasmodium falciparum mRNA for major surface antigens precursor (P195).		
ACCESSION	X02919.1	GI:9864	
VERSION	X02919.1		
KEYWORDS	antigen; direct repeat; signal peptide; surface antigen.		
SOURCE	Plasmodium falciparum.		
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE	1 (bases 1 to 5917)		
AUTHORS	Holder,A.A., Lockyer,M.J., Odink,K.G., Sandhu,J.S., Riveros-Moreno,Y., Nicholls,S.C., Hillman,Y., Davey,L.S., Tizard,M.L.V., Schwarz,R.T. and Freeman,R.R.		
TITLE	Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum merozoites		
JOURNAL	Nature 317 (6034), 270-273 (1985)		
MEDLINE	86014355		
PUBMED	2995820		
REFERENCE	2 (bases 1 to 5917)		
AUTHORS	Holder,A.A.		
JOURNAL	Direct Submission		
TITLE	Submitted (06-MAR-1991) Holder A.A., National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK		
REMARK	Update of published sequence		
COMMENT	Data kindly reviewed (06-MAR-1986) by R.R. Freeman.		
FEATURES	Location/Qualifiers		
source	1..5917		
CDS	/organism="Plasmodium falciparum"		
	/db_xref="taxon:5833"		
	418..5337		
	/note="P195 precursor"		
	/codon_start=1		
	/protein_id="CAA26676.1"		
	/db_xref="GI:9865"		
	/db_xref="Swiss-Prot:P04933"		
	/translation="MKTFPCSLRFILPIINOCYTHESYOYLKFLALDVAITGSV LPQEKMYLNGSTGLAVTTTSPESSKSVASGGSGYSVASGVSAGSVASGSVAS GGSSNRKTNPSSSDSAKSYADLKHVRANKVLITREKLIPOLFUTNMHLLCON IHGRYLIDGYEELNELLKYLFEPDFDLRALNDNCANDQCQIFPNKIIRANELDYIK KLIVGRRPDNIKNQVNGKMDYFKKKTKTENENLENEIEBCKTIQNKNAKKEERK KLVOAYDI.SLYNKOLEBAHLISVLRIDOTLKNNIKELDKIINEIKVPSPANGS NTFPLLDKNKITEHHEKETIKELAKTIKFNDLS.FTPPLELEYLRKKNNIDISAVV ETKSTEPNETPGCVTPLESTINDLNMLNLNLSGDLINPDVTKEPSKNTYDNERR KFINELKRIKERKLESDEKSYEDSKSLINDTYKEKLIENIYDSKTNMNDLTNN PEKMGRKRSYKVEKLHTHNPFASYENSKNLEKLTALKMEDYSLRNVIVEKELY YKNLSIKIENEIEFTLVENIKKDEOLEEKTITDENRDPDEKILEVDIVAOVKVLL MNLDELKTOLILKANVELKHHHPNSVSKOENKOOPYLYLLEKLEDKLVMPRYVE SLIIEEKNITKGOSONSSEPSTGEITGOATTTCPGOAGSLAEGDSVQAQABOKRA OPPLPVPPPEKKAQVPTPPPAVNKKTEWNSLDTLEKYLEPLINTSYCIKYLIVSHT MNEKLQYKITTKEERSLSCDPLDLFIPLQNQNI.PWYSFDSLNSLSLOLEWEITE KEWNCVNIKLDNDIKRLNELKAVSTSVYTLSSSNQPLSLTPDRKPYASNDPTS HSTNLNMSLKFEFNILSIGKRNIYIOELIGKSSENYEYLIKDSDFYNSESFPNFVK		

	sig_peptide	418..474 /note="putative"
	mat_peptide	475..5334 /product="region coding for 83k, 42k and 19k major surface antigens (aa 1-1628)"
	repeat_region	631..728 /note="alternating repeated SVA-SCG (aa 72-104)"
	misc_feature	763..765 /note="pot. glycosylation site"
	misc_feature	1219..1221 /note="pot. glycosylation site"
	misc_feature	2707..2709 /note="pot. glycosylation site"
	misc_feature	2719..2721 /note="pot. glycosylation site"
	misc_feature	2764..2766 /note="pot. glycosylation site"
	misc_feature	2947..2949 /note="pot. glycosylation site"
	misc_feature	3175..3177 /note="pot. glycosylation site"
	misc_feature	3307..3309 /note="pot. glycosylation site"
	misc_feature	3589..3591 /note="pot. glycosylation site"
	misc_feature	3910..3912 /note="pot. glycosylation site"
	misc_feature	3937..3939 /note="pot. glycosylation site"
	misc_feature	4750..4752 /note="pot. glycosylation site"
	misc_feature	4750..4752 /note="pot. glycosylation site"
	BASE COUNT	2644 a 636 c 727 g 1910 t
	ORIGIN	
	Query Match	57.0%; Score 2816.4; DB 3; Length 5917;
	Best Local Similarity	73.2%; Pred. No. 0;
	Matches 3609;	Conservative 0; Mismatches 1321; Indels 0; Gaps 0;
OY	10 ATGAAATCATTTTCTCTCTCTGTCATTTTCGTTTTTATCATCATATCATCGTCGTG	69
DB	418 ATGAAGATCATATTCCTTTTATGTCATTTCTTTTATATTAATATACAAATGTGTA	477
OY	70 ACCCAGATCCTATCAGAGAGCTGTGTTAAAGAACTGAGAGCTTTGGAAGATGCCGCTCT	129
DB	478 ACACATGAAGATATCAAGAACTTGTCAAAAAAGTAAAGCTTTGAAGATGCACTATTTG	537
OY	130 ACCGATACAGCGCTTCCAGAGAGAAGATGCTGCTGAATGAAGGGACAGAGTGCACAG	189
DB	538 ACAGCTTATAGTTTATTTTCAAAAAGAAAAAAGTATTAAAGTAAAGAGAACCAAGTGAACA	597
OY	130 GCCGTTAACACAGACACCGCGTCTTAAGAGGTGTGGCTGAGCGAGGGGCGTCCGCTGGG	249
DB	598 GCTGTTAACACAGTACACCTGTGCTTCAAGAGGTTTATGTTGCTTCAAGTGCTTCAAGTGGC	657
OY	250 TCTGTGGCCTCTGGGGGTTCGCTCGCTCCGCGGACGCTGAGTCATCAGTGGCTCAGTG	309
DB	658 TCAAGTGTCTCAGTGGGCTCAAGTTCCTTCAGGTGCTCAGTTGCTTCAAGTGGCTCAGTT	717
OY	310 GCAGCGCGGCTTCGGGACAGTGAAGAACCAATTCATCTGACAACTTAAGCATTC	369
DB	718 GCTTACAGGTGGTTCAGGTAATTCAGAGAGTCAAAATCTTCAGATTAATTCAGTGATTTCA	777

Oy	370	GAGCCCAAGTCCTAAGCCGACCTCAGACACCGAGTGAAGAACTATCTCCTACATATCAAG	429
Db	778	GATCCTAAATCTTACGCTGTGATTTAAACACAGAGTACGAAATTAACCTTTATCTTAACTATCAAA	837
Oy	430	GAGCTGAGTACCCACAGTTGTTCGACCTCCACTCAATCATATGCTGACACTGTGTATAC	489
Db	838	GAACTCAAAATATCCTCAACCTTTTGATTTTAACTAATCATATCTTAACCTTTGTGTATAT	897
Oy	490	ATTATGCGCTTCAAAATATCTGATTGACGCTTACGAAGATCATGAACCTCCTGTACAG	549
Db	898	ATTATGCTTCAAAATTTTAAATTTGATGATATGAAGAAATTAAGAATTTATATATAA	957
Oy	550	TTGAAATTCCTACTCGACTGCTTAAGGGCAACGAATGAGCTTTGGGCCAATGACAT	609
Db	958	TTAAACTTTTATTTTGATTTTATTAAAGCAAAATTAATGATGTATGTGCTAATGATTTAT	1017
Oy	610	TGTCAAAATCCATTCAAATTTGAAGATCAGACCCAAACGAGTTGGACGTATTAAGAAGTTG	669
Db	1018	TGTCAAATACCTTTCATCTTAAATTCGTGCAAAATGAATTTAGACGTACTTAAAAAACTT	1077
Oy	670	GTCCTGGATATGTGCAAGCCTCTCGACACATCAAGCAATATGTCGGCAACATGGAAGAT	729
Db	1078	GTTGTGGATATGAAAAACCATTAGCAATATTTAAAGATTAATGTAGAAAAATGGAAGAT	1137
Oy	730	TATATTTAAAAAGATTAAGAAAGACCATCGAAGACATTTAACGAGCTATGGAAGATCCAAA	789
Db	1138	TACATTTAAAAAATTAAAAAACCATAGAAATATTAATGAATTAATTTGAAGAAAGTAAAG	1197
Oy	790	AAGACCATAGCAAAATTAAGATTCGACCAACGAGGAAGAAAGAGACAGTGTACCAG	849
Db	1198	AAAACAATTTATAAAAATAAGATGCACTAAAGAAAGAAAAAATTTATACCAA	1257
Oy	850	GCCAGTAGACCTGTCCATCTTAACTAAACAGCTTGAAGAGCCCTATACCTCATACAG	909
Db	1258	GCTCAATATGATCTTCTCTATTTCATTAACATTAAGAAAGACATTAATTTATATAGC	1317
Oy	910	GTACTGAGAAAGCGCATGACACCTCAAGAAAGATGAAATATCAAGAACTGCTCGAC	969
Db	1318	GTTTAAAGAAAAAGTATTCACACTTTAAAAAATAAGAAATTAAGGAATTTACTGAT	1377
Oy	970	AAGATTTAATGAATTTAAGAAATCCTCCGCGACCACTCGGAGACACCCCTATACACGCTG	1029
Db	1378	AAGTTAATTAATTAATTAATAAATCCCGACCGCCAAATTCGGAATTAACCAAAATCTCTC	1437
Oy	1030	CTGACACAGAACAAGATAGAGAGACGACGAGAAAGAGATCAAGAAGATGCCAAAAC	1089
Db	1438	CTTGATTAAGAACAAAAATCGAGGAAACACGAAAAAGAAATTAAGAATTTGCCAAAAC	1497
Oy	1090	ATTAGTTCAACATAGATTCTCTTTACTGATCCCTGTGAGCTGGAGTACTACTTGAGA	1149
Db	1498	ATTAAATTTTATATGATAGTTTATTTACTGTGATCCACTGGAATTTGAATTAATTTAAGA	1557
Oy	1150	GAGAAAGATTAAGATTTAGACATCTCCGCAAGTCGAGACAAAGAGATCAACGACCT	1209
Db	1558	GAAAAAATTAATTAATTTGATTAATAGTGCAAAAGTTTGAACAAGAGATCAACTGAAACC	1617
Oy	1210	AATGAATATCCCAATGCTGTGACGTAACCCCTGTCTTAATACGATATCACACGCTCTC	1269
Db	1618	AATGAATATCCCAATGAGAGTACTTACTCTTGTCAATTAACGATATTAACAATGCTTTA	1677
Oy	1270	AAGAGCTCAATGCTGTGCTGACTGATTAAACCCCTCGATTAATACAAAGAACCTCT	1329
Db	1678	AATGAACCTTAATCTTTTGCTGTGATTTAATTAATTCATTTGATTAATCAAAAGAACCAAG	1737
Oy	1330	AAGAAATTCACACAGACATAGAGAAAGAGTTTATCAACGAATCAAGAGACATC	1389
Db	1738	AAAAACATATATCTGTATATGAAAGAAAAAATTCATTAATGAATTAAGAAAAAAT	1797
Oy	1390	AAAATTAAGAAGAAATTTGAGAGTGCACAAGAAAGTTACGAAGACCGACGCAAAAGT	1449
Db	1798	AAAATTAAGAAAAAATTTGAAATTTGATTTAAAAAATCTTTACGAAGACAGATTTCTAAGTCT	1857

QY	1450	CTAAACGATATCACTAAAGAGTATGAAAAGCTCTCAAGAGATCTATGATTCGAAATTC	1509
Db	1858	TTAAATGATATACAAAAGAAATGAAAAATTCTTAAGAAATTTATGATAGCAAAATTC	1917
QY	1510	AACATTAACATCGAGCTGACCACCTTCGAGAAAAATGATGGAAAAAGGACTCTTACAAA	1569
Db	1918	AATATATATATAGATTTTAACTAATTTTCGAAAAATATGAGTAAAGATATTCTATATAA	1977
QY	1570	GTGGAGAAACTGACACACCAATATACCTTTCATCTATAGAAATTTCTAAGCATATCTT	1629
Db	1978	GTTCGAAACCTTACACACCAATATACCTTTGCATCTTATGAAAAATTTCTAACATATCTT	2037
QY	1630	GAGAAAGCTACCAAAGCTCTTAAAGTATATGAGAGCATATTCTCTCGGAACTTTTGTG	1689
Db	2038	GAAAAATTAAACAAAAGCTCTTAAATATATGGAAGATTTATCTTTAAGCAATATACTACTT	2097
QY	1690	GAGAAAGAACTAAAGTATTTCAAGAATCTCATATAGATACATGGAAGAGAGATCGACAGC	1749
Db	2098	GAAAAAGATTTAAATTTTATTAATAATTTTATATAGCAAAATATGAAAATGAGATTGAACA	2157
QY	1750	CTTTGTGGAACATTAAAGAGATGGAAGAACACTGTTTGTGAGAGAGATTTACAAAAGC	1809
Db	2158	TTAGTTGAAATATTTAAAAAAGATGAAGAACACTTTTGTGAAAAAAAATTTACTAAAGC	2217
QY	1810	GAAAAATTAACCAAGTGAAGAAATCTCTGAGAGTCTCCGATATTGTTTAAAGTCCAAAGTCAG	1869
Db	2218	GAAAAATTAACCAAGTGAAGAAATTTTATGAAAGTATCTGCAATTTGTAAAGTCAAAAGTCA	2277
QY	1870	AAGGTCCTCATGAAACAAAGATTGATGAACTCAACACATCACTCACTCTGTGAAGAC	1929
Db	2278	AAAGTTTATTAATGAACAAAAATTTGACGAATTAATAAAAGACTCAATGTATTTAAAAAT	2337
QY	1930	GTGAGATTAAACCAATATATATACATGTGCCGAATAGTTATTAACAGAGAAATTAACAGAA	1989
Db	2338	GTAGAAATTAACCAATATATATACATGTGCCAATCTTTACAAACAAAGAAATATAGCAAGAA	2397
QY	1990	CCATATCTACTCATCTGTAAGTCAAGAAAGAGATAGCAAACTGAAAGTGTCTATGCTCCAAA	2049
Db	2398	CCTTATTTATTTAATTTGTGTTGAAAAAAGAAATGTATTAATTTAAAGTGTCTATGCTCAAG	2457
QY	2050	GTCCAGAGCCGTGATCAACGCAAGGAAGAGAAGAAACATTAATAACGGAAGAGCTCAAGATAC	2109
Db	2458	GTAGATCATTTGATTAATTAATGAAGAAAAAAAACATTAATAACGGAAGGTCAATCGGATAT	2517
QY	2110	TCCGAGCCCTTCACAGAGAAGAGATATACCCGACAGCTTACCCACCAACCCGAGCAACAG	2169
Db	2518	TCCGAAACCATTAACCCGAAAGGAATTAACAGCAACAGCAACTTACAAAACCTTGACACAA	2577
QY	2170	GCCGGTTCAGCTCTCGAAGCCGATAGCTGCAACGTCACAGCAACAGACAGAAAGCAGCA	2229
Db	2578	GCAGATCTGCTTTAGAAAGAGATTCAGTCAAGCAACAGCAAGCAAGAAACAAAGCA	2637
QY	2230	CAGCTTCAGTGGCCAGTGGCCGTTCCAGAGAGCTAAAGCTCAAGTGCCTTACACACACAGCT	2289
Db	2638	CAACACACAGTACAGTACCAAGTACCAAGAAAGCAAAAGCAAGTCCCAACACACACAGCA	2697
QY	2290	CCTGTGAATTAACAAGACCGAAGTGCAGCAAACTGGACTACTCTGAGAGAGCTGTATGAG	2349
Db	2698	CCAATTAATTAATTAACCTGAAAATGTGTTCCAAATTAATGATTAATCTTGAAAAAATTAATGAA	2757
QY	2350	TTCTCGAATACATCTCAATCTGCAACCAATATATCTCTGCTCTCACAGCAGCATATGAC	2409
Db	2758	TTTTTAATTAATCTCATATATATATGTCACAAATATATTTTGTTTCAACCTCAACATATGAC	2817
QY	2410	GAGAGATTTCTTAACACGTCAAGATTAACAAAGAAAGAGAGATTAACCTGTCTTGT	2469
Db	2818	GAAAGATTAATTAACCAATTAATAATTAACAAAGAGAGAAAGCAAAATTAAGTTCATGT	2877
QY	2470	GATCACAGGACCGGTTCATATTCAGAACACACATTCGCCGTATATATTTATGATGTC	2529
Db	2878	GATCATATTAGCTTATTTATTAATTAACAAATTAACATACCTGTATGATGATTTCTATGTTT	2937
QY	2530	GATGACCTCAACATTTCTCTCTCAACGTTTCATGAGAAATTAATGAAGAGAGATGCTC	2589

DB 2938 GATAGCTTAACAAGTATGTTATTCACAACTATTATGGAATTTATGAAAAAGAAATGTT 2997
QY 2590 TCGAACCTGATATACACGAAGACAGACAAATTAAGAACTCTCGGAGGAAGCTAG 2649
DB 2998 TGTAAATTTATATTAACCTTAAGATTAAGATAAATAAATTTATTAAGGAAGCCAAA 3057
QY 2650 AAGGTCTCCACTCTGTTAAACTCTCTCCAGCTCCATGCAACACAGCTCTCTGACA 2709
DB 3058 AAGGTATCCACTCTGTTAAACTCTCTCCAGCTCCATGCAACACAGCTCTCTGACA 3117
QY 2710 CCTCAGACAGACCCGAGAGTGAAGCTACAGACACCTCTCACTGACCAACTTAAT 2769
DB 3118 CCTCAGAGTAAACCCGAGAGTGAAGCTACAGACACCTCTCACTGACCAACTTAAT 3177
QY 2770 AACTCACTGAAAGCTGTTGGAACATCCCTCTCGGAGAGATATAGAATCATCTACCA 2829
DB 3178 AATAGATTTAAATATTTATTAAGAAACATTAAGATCTTGGAATAAATAAATATATACCA 3237
QY 2830 GAACCTATTGGAGAGAAATCGTCGAGAACTCTACGAGAGATTAAGAAAGAGCGAC 2889
DB 3238 GAATTAATAGTCAAAAATAAGTGAATACTTTATGAAAGATATTAAGAAAGATGAT 3297
QY 2890 ACATTTCTATACGAGAGCTTCACTACTCTGTAATCTTAAGCCGATGATATCACTCT 2949
DB 3298 ACATTTTATATGATTAATCTTTTACAAATTTGTAAATCTTAAGCTGATGATTAATCA 3357
QY 2950 CTTAAGCATGATCTTAACCTAAGAACCTGAGAGACATCAATAGCGAAGAGACA 3009
DB 3358 TTGAATGATTAATCAAAAAGAGAAATTAAGAAAGATTTAATTAATTAATTAATTAAT 3417
QY 3010 CTGCAACTGAGCTTCGACTCTGACAAACAGTACAACTGAGAGACTCTTCGAC 3069
DB 3418 TTACAGTTATCATTTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTTGAT 3477
QY 3070 AAAAGAGAGACATCGGCAAGTATAGATGCAATCAAGAACTGACTCTGCAAGAG 3129
DB 3478 AAAAGAGAGACATCGGCAAGTATAGATGCAATCAAGAACTGACTCTGCAAGAG 3537
QY 3130 CAGCTTGAAGAGCAACTCAACTGCAAGCAATCGAAGACAGTACTGCAAGACTCTCA 3189
DB 3538 CAGTTGAATCAAAATTTGAATTCATTAATACCAAGATGATTAACAACTTTCT 3597
QY 3190 GTGTTCTTCAAGAGAAAGAGCCGAGATGCGAGACAGAGAACTCTGAGAGAC 3249
DB 3598 GTTCTTCTTCAAGAGAAAGAGCCGAGATGCGAGAACTGCAAACTGATTAAGAAAC 3657
QY 3250 ACCAAGTCTTCTCAAGACATCAAGGCTGCAAGTATTAATGAGGAGCTCTCT 3309
DB 3658 ACAAAATTTATTTAGACATTTAAGAGACTGTTAATTAATTAATTAATTAATTAATCT 3717
QY 3310 CCTCTGAAGACTCTCTCCGAGAGAGATCCAGACCGAGAGATTAAGCTCCGAGCTCTGAG 3369
DB 3718 CCATTAATAAATTTAAGTGAAGATCAATTCAAACAGAAATTAATTAATTAATTAATTA 3777
QY 3370 AACTTCAAGGTCTGCTTAAGCTCGAAGCAAGCTGAGAGCAACTGAACTGAGAGAG 3429
DB 3778 AACTTCAAGGTCTGCTTAAGCTCGAAGCAAGCTGAGAGCAACTGAACTGAGAGAG 3837
QY 3430 AAGAAGCTGAGTACCTCTCTAGCGAGCTGATCACTGATCCCGAGCTCAAGAGAGTC 3489
DB 3838 AAAAATTTATCATCTATTCATCAAGTGAATCAATTAATTTCTGATTAATAAGAGTA 3897
QY 3490 ATTAAGACAAGAACTACAGCGCATAGCCCAAGAGAGAGATTAATACAGCGTAATAC 3549
DB 3898 ATTAAGACAAGAACTACAGCGCATAGCCCAAGAGAGAGATTAATACAGCGTAATAC 3957
QY 3550 GCAGTGAATCTTACAGAGAGTCTGCTGAGAGAAACAGATGCTGCACTGAGTCT 3609
DB 3958 GCATTTGAATCTTACAGAGAGTCTGCTGAGAGAAACAGATGCTGCACTGAGTCT 4017
QY 3610 GAATCTGCTCGACACTGAGAGAGTCTCAACCTAAGAGCTGATCTACTCATGTC 3669

DB 4018 GAAAGTGATCCGACACATTAAGAAACAAGTCAACCAAGAAACAGCATCACTCATGTA 4077
QY 3670 GGAGCGAGTCCAAATCAATTAACCATCTCAGACGCTGAGAGTGAAGTGAAGCTGTC 3729
DB 4078 GGAGCGAGTCCAAATCAATTAACCATCTCAGACGCTGAGAGTGAAGTGAAGTGAAGCTGTC 4137
QY 3730 ATCATGCTGCTTATGCTGCGAGAGAGAGAGAGTACAGTCACTGCGAGTGTGTC 3789
DB 4138 ATCATGCTGCTTATGCTGCGAGAGAGAGAGTACAGTCACTGCGAGTGTGTC 4197
QY 3790 ACCGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3849
DB 4198 ACAGAGAAACCTTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4257
QY 3850 TACAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3909
DB 4258 TATGAGTCTTATATTAATTAACCTTTAGAGAGTGTATAGAGTTTAAAAACAATTA 4317
QY 3910 GAGATTAACGTGATGACCTTCAATGTCACAGTGAAGAGATTTGACAGCCCTTTAAT 3969
DB 4318 GAAATTAACGTGATGACCTTCAATGTCACAGTGAAGAGATTTGACAGCCCTTTAAT 4377
QY 3970 AAGAGAAATTTCAAGAGAGTCTGAGAGAGTCTGAGAGAGTCTGAGAGAGTCTGAGAG 4029
DB 4378 AAGAGAAATTTCAAGAGAGTCTGAGAGAGTCTGAGAGAGTCTGAGAGAGTCTGAGAG 4437
QY 4030 TCCCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4089
DB 4438 TCAAGTATTAATTTGCTGCAAGATCCATTAATTTCTTATTAAGAAAAAGAGATAA 4497
QY 4090 TTTCTGCTGATTAACATTAATTAACAGTCTGATGAGAGTCTGATGAGAGTCTGATGAG 4149
DB 4498 TTTCTGCTGATTAACATTAATTAACAGTCTGATGAGAGTCTGATGAGAGTCTGATGAG 4557
QY 4150 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4209
DB 4558 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4617
QY 4210 AAAAGATTAATCAAGATTAAGAGAGGAGATTAAGATTAAGATTAAGATTAAGATTA 4269
DB 4618 AAAAGATTAATCAAGATTAAGAGAGGAGATTAAGATTAAGATTAAGATTAAGATTA 4677
QY 4270 ATGGAACCTGTAACAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAG 4329
DB 4678 ATGGAACCTGTAACAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAG 4737
QY 4330 GCCAAGTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4389
DB 4738 GCCAAGTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4797
QY 4390 AACTACCTCAAAACATCAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAG 4449
DB 4798 AACTACCTCAAAACATCAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAG 4857
QY 4450 GGAATGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4809
DB 4858 GGAATGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4917
QY 4510 ACTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4569
DB 4918 ACTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4977
QY 4570 CTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4629
DB 4978 CTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5037
QY 4630 AGCGCTGCTTCAAGCATCTGCAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGA 4689
DB 5038 TCGATGATGCTTCAAGCATCTGCAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGA 5097
QY 4690 CAAGAGAGATTAAGTGTGCTGAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGA 4749
DB 5098 CAAGAGAGTGAATGATGCTGAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGA 5157

[illegible][illegible]

OY	850	GCCAGTGCACCTGTCCATCTATTAACAACAGCTTGAAGAAGCCCATPACCTCATCAGC	909
Db	2432	GCTCAATATGATCTCTTCTTATTTACATTAACAAATTTGAAGAAGCACATTAATTTATTAAGC	2491
OY	910	GTCACGGGAAGGCCCATFAGACACCCCTCAAGAAGTAATGAATAATATCAAGAACTGGTCAGC	969
Db	2492	GTTTATTAGAAAACGTAATTGCACCTTTTAAAAAAAATGAAAACATTTAAGAAATTTACTTGAT	2551
OY	970	AAGATTAAATGAATTAAGAATTCCTCCGACAGCCACTCGGGAGACCCCTAACACGCTG	1029
Db	2552	AAGATTAATGAATTAATAAATTCGCCACCGGCCCAATTCGAAAAATACACAAATACTCTC	2611
OY	1030	CTGGACAAGAACAAAGAATFAGAGAGACACGAAAGAAATCAAGAAATGCCAAACC	1089
Db	2612	CTTGATTAAGAACAAAAAATCGAGAAACGAAAAAGAAATTAAGAAATTTGCCAAACT	2671
OY	1090	ATTAAAGTTCAACATAGATTCTCTCTTACTAGATCCCTTGACCTGGAGTACTCTTGGA	1149
Db	2672	ATTAAATTTAATATGATTAATTTATTTACTGATCCACTTGATTAAGATTAAGATTTAGA	2731
OY	1150	GAGAAATTAAGAAATATAGACATCTCCGCAAGTCGAGACAAGAAATCAACGACCT	1209
Db	2732	GAAAAAAATATAAATATTGATATAGGCAAGAGTTGAACAAAGAAATCAACGTACACC	2791
OY	1210	AATGAATATCCCAATGGGTGAGCGPCCGCTGTCTTAAGAGATPACAAACGCTCTC	1269
Db	2792	AATGAATATCCCAATGGAGATTACTTATCTTGTCTCATFAGAGATTTAACAATGCTTTA	2851
OY	1270	AACGAGCTCAATAGCTTCGGTGCATTTAAACCCCTTGATTAAGCAAGAACCCCTCT	1329
Db	2852	AATGACTTAATTTCTTTTGGTGAATTAATTAATCCATTTGATTTATACAAAGAACCAAGT	2911
OY	1330	AAGATATCTACACAGCATTGAGAGAAAGATTATACAAGAAATCAAGSAGAAAGTC	1389
Db	2912	AAAAACATATATCTGTAATFAGAAAGAAAAAATTCATAAAGAAATTAAGAAAAAATTT	2971
OY	1390	AAAAATGGAAGAAGAAATTTGAGAGTGCACAAGAAAGTTACGAAGACCGCAGCAAAAGT	1449
Db	2972	AAAAATAGAAAAAATAAATTTGATCTGATFMAAAATCTTCAGAGACAGATCTTAAGCT	3031
OY	1450	CTAAACGATATCACTAAAGATFAGAAAAGTCTGTGACAGAGATCTATGATTTCCAAATTC	1509
Db	3032	TTAAATGATATTAACAAAAGAAATFAGAAAAATTTCTTAAGAAATTTATGATAGCAAAATTC	3091
OY	1510	AACATATAACATCGACCTGACCAACTCGSAAAAATGATGGSAAAAAGTACTCTTACAA	1569
Db	3092	AATATATATATAGATTTTAACCTAATTTTCGAAAAATGATGGTAAAGATATTCATATAA	3151
OY	1570	GTGGAAGAACTGACACACCATTAATACCTTTGCATCTATGAGAAATTCATAGCATATCTT	1629
Db	3152	GTTGGAAGAACTTACACACCATTAATACCTTTGCATCTATGAGAAATTCATAGCATATCTT	3211
OY	1630	GAGAGACTCACCAAGCTCTTAAAGTATFAGSAGACTATTTCTCGCGAACAATTTGTGTG	1689
Db	3212	GAAGAATTAACAAAAGCTCTTAAATATFAGSAGATTAATTTCTTAAAGSAAATATAGTACT	3271
OY	1690	GAGAAAGAACTAAAGCTTTCAAGAAATTCATFAGTAACATGGAAGAAAGAGATGAGACG	1749
Db	3272	GAAGAAGAAATTAATTTATTAATAAATTTATATAGCAAAATFAGAAAAATFAGATTTAAACA	3331
OY	1750	CTTGTGGAACATTTAAGAGGATFAGAAAGACAGTTGTTGAGAGAGAAATTTACAAAGAC	1809
Db	3332	TTAGTTGAAATATTTAAAAAAGATFAGAAAGACACCTTTTGAAAAAATAATTACTAAAC	3391
OY	1810	GAATAATTAACAGATFAGAAAGATCTGAGAGTCTCCGATATTTGTTAAATCCAAAGTCAG	1869
Db	3392	GAATAATTAACAGATFAGAAAGATTTTAAAGAGTATCTGACATTTGTTAAAGTACAAAGTCCA	3451
OY	1870	AAGGTGCTCCCATFAGAACAGATTTGAATGAACCTCAACACATCACTACTTCTGAAAGAC	1929
Db	3452	AAAGTTTATTAAGAACAAATTTGACGAATTTAAAAAAGACATCAATGATTTTAAAAAAT	3511

QY	1930	GTGAGTAAACAACTATATACACTGTCGCGAATAGTTATATACACAGAGATTAAGACGAA	1989
QY	1990	CCATACCTACCTCATCTGATCTCAAGAAAGAGATAGCAAACTAAAGTGTCTATGCCAAA	2049
QY	2050	GTCCGAGACCTGATCAACGAGAGAGAGAACATTAAAACTGAAAGACGTCAGATTAC	2109
QY	2110	TCCGAGCGCTCCACAGAGAGAGATATACCGGACAGCTACACACAGCGCCGAGCAACAG	2169
QY	2170	GCCGCTCAGCTCCGAAAGCGATAGCGTCAAGCTCAAGCAACAGACAGAGACGAGCA	2229
QY	2230	CAGCCTCAGTGCAGTGCCTGTTCCAGAGGCTAAAGCTCAAGTCCCTACACACAGCT	2289
QY	2290	CCTGGAATTAACAGACCGAGATGTGACAGAACTGACCTACTCTGAGAGTGTATAG	2349
QY	2350	TTCTCGAATACATCTCATCTGCGCAAAATATATCTCTGCTCCACAGCATATGAC	2409
QY	2410	GAGAAGATTCTTAACAGTCAAGATATACCAGAGAGAGAGATAACTGCTCTTGT	2469
QY	2470	GATCACGACGCGCGCTCAATATCCAGAACAACTCCGTTATGATTTATCTATGTC	2529
QY	2530	GATGAGCTCAACATTTCTCTCAACTGTTCCATGGAATATATGAGAGAGATGCTC	2589
QY	2590	TGCAACCTGTAAACCTCAAAAGCAACGACAGATTAAAGACCTTCTGAGGAGCTAAG	2649
QY	2650	AAGGCTCCACCTGTTAAACTCTCTCTCCAGCTCATGCAACACACTGTCTCTGACA	2709
QY	2710	CCTCAAGACAAGCCCGAAGGACGCTTAACGACGACACTCTCACTGACCAACTTAT	2769
QY	2770	AACTCACTGAACCTGTTGGAAACATCTGCTCTCGGCAAGATTAAGACCTTACCA	2829
QY	2830	GAACTTATGACAGAAATGCTCCGAGAACTTCCAGCAAAATATCTGAAGACGACAC	2889
QY	2890	ACATTCTATAACGAGACTCTCACTACTCTGTAATCTTAAGCCGATGATATCAACT	2949
QY	2950	CTTAACGATGATTAACCTTAAGACGTCGAGAGACATCAATTAAGCTGAAGACGA	3009
QY	3010	CTGACACTGAGCTCGACTGTACAAACAGTACAACTGAAACTGAGAGACTTTCGAC	3069
DB	3512	GTAGATTTAAAACTAATATATACATGTTCCCAATCTTCAACAAAGAAAAATTAAGACAA	3571
DB	3572	CCTTATTATTAAATGTGTGAAAAAAGAAATTGATTAATTTAAAGTGTCTATGCTCAG	3631
DB	3632	GTAGAATCATGATGAATTAATGAGAAAAAAAACATAAAAACGAAAGTCAATCGGATT	3691
DB	3692	TCCGAGCGCTCCACAGAGAGAGATATACCGGACAGCTACACACAGCGCCGAGCAAC	3751
DB	3752	GCAGGATCTGCTTAAAGAGAGATTCAGTATCAAGCAACAGCAACAGAAACAAAAACA	3811
DB	3812	CAACACACAGTACAGTACCAAGTACCAAGAAAGCAAAAGCAAAAGTCCCAACACACACA	3871
DB	3872	CCAGTAATATTAACCTGAAATGTTCCAAATTGATTAATCTTGGAAAAATTAATATGA	3931
DB	3932	TTTTTAATACTTCATATATATGTCACAAATATATTGTTGTTCACTCACTATGAC	3991
DB	3992	GAAAAAGATTTAAACAACTTAAATTTACAAAGAGAGAAAGCAAAATTAACTGACAT	4051
DB	4052	GATCACGACGCGCGCTCAATATCCAGAACAACTCCGTTATGATTTATCTATGTC	4111
DB	4112	GATGAGCTCAACATTTCTCTCAACTGTTCCATGGAATATATGAGAGAGATGCTC	4171
DB	4172	TGTAATTTATTAACCTTAAGATTAATGACAAATTTAAATTTATTAAGAGAGACGAA	4231
DB	4232	AAAGTCACTCAATTTCTCTCAACTGTTCCATGGAATATATGAGAGAGATGCTC	4291
DB	4292	CCTCAAGACAAGCCCGAAGGACGCTTAACGACGACACTCTCACTGACCAACTTAT	4351
DB	4352	AATGATTTAAAAATTTATGAAACATATGAGCTTCGAAAAAACAAAAATTTATACAA	4411
DB	4412	GAAATTAATAGTCAAAAACTAGTGAACCTTTATGAAAAAGATTTTAAAGATGTAT	4471
DB	4472	ACATTCTATAACGAGACTCTCACTACTCTGTAATCTTAAGCCGATGATATCAACT	4531
DB	4532	CTTAACGATGATTAACCTTAAGACGTCGAGAGACATCAATTAAGCTGAAGACGA	4591
DB	4592	TTGGAATGATGATTAAGAGAGAAATTTAGAAAGAGATTAATTAATTTAAAAAAACT	4651
DB	4652	CTGACACTGAGCTCGACTGTACAAACAGTACAACTGAAACTGAGAGACTTTCGAC	4711

[illegible]

D	b	5672	TTCCTTAAGCAGTTAAATATTATTTAAGCATTCATAGATACGGATATATAATTTTGCAAAT	5731
O	y	4150	GATGCGTGCGGGGTATTATACAAGATCCGTGAACGAAAAATACAACTCTGACCCTTGACTCTATT	4209
D	b	5732	GATGTTCTTGGAATATTATPAAATATTATTCGCAAAAATATPAATCAGATTTTGATTCAAATT	5791
O	y	4210	AAAAAGTATATCACGATPAGCAGGCGAGAATGAAAAATATCTGCCCTTCTCGATATAC	4269
D	b	5792	AAAAATATATTCACGACAAACAAAGTGAAAAATGAAAATACCTCCCTTTTAAACAAT	5851
O	y	4270	AATGCAAAOCCTGTACAAAGCATGGAACGAAATATCGACCTCTGTGTAATTCACCTGGAG	4329
D	b	5852	ATTGAGACCTTATATPAAACAGTTATATGATATTAATTTGATTTATTTGTAAATTCATTAGAA	5911
O	y	4330	GCCAGAGTCTCTCACTATCTTACGAGAGAAGCAATGTGGAAGTTAAATCAAGAGAGCTG	4389
D	b	5912	GCAAAAGTTCTAAATTTATACATATGAGAAATCAAACGTAGAAAGTTAAATTAAGAAGACTT	5971
O	y	4390	AACTACTCTAAACAATCCAAGACAAAGCTGGCAGATTTTCAGAAAAATPACATTTTGGTC	4449
D	b	5972	AATTTACTTAAAAAACATTCACACAAATTTGGCAGATTTTAAAAAAAATPACATTTTGGTC	6031
O	y	4450	GGAATTTGACAGCTGTCTACGATATATACCAACAACATTCCTCCAGCAAGTTCTGTGCC	4509
D	b	6032	GGAATTTGATGATTTATPACAAGATTTATACCAATAPACTTATTTACCAAGATTCCTTAGT	6091
O	y	4510	ACTGCACTAGTGTGTGAAAAACCTCCGCCAAACAGTGTGAGACAAATCTCGACGCGCAAC	4569
D	b	6092	ACAGGTATGTTTTTGAAAAATCTTGCTAAAAACGTTTTATCTAATTTTACTTGATGGAAGAC	6151
O	y	4570	CTGCAAGGCGATCTGACATCTCCAGCACCAATGCTGGAAGAAACAGTGCCTCCAGAT	4629
D	b	6152	TTGCAAGGATGTATTAACATTTTCACACACCAATGCGTAAAAAAAACAATGTCCACAAAT	6211
O	y	4630	ACCGCGCTGTTTACGATCTGAGACAGGCGCAAGAGTGCAAGTGCTGCTCTGTAACATCAAA	4689
D	b	6212	TCTGATGTGTGACGATTTAGATATGAAGAAGATGTAAATGTGTTATTTAAATTTCAAA	6271
O	y	4690	CAGAAGAGATAACTGCGTGGAGAAACCCAAACCTTACTCGAATGAANAACAANTGCGGG	4749
D	b	6272	CAAGAAGGATAAATATGTGTGAANAATCCAAATCTTACTGTGAAGAAAATATGTTGGA	6331
O	y	4750	TGTGACGCCGATGCTAAATGCAACGAGAGACAGCGGCTTAACGGAAGAAAATCACA	4809
D	b	6332	TGTGATGCGAGATGCCAAATGTACCCNAGAAGATTCAGGTAGACGAGGAAGAAAATCACA	6391
O	y	4810	TTCGAGTGTACTAAACCCGACACTCTACATCTGACGAGGATTTTGTGCCGACGCT	4869
D	b	6392	TGTGATGTACTAAACCTGATTTCTTATCCACTTTCAGATGATGATTTTCTGCTCCCT	6451
O	y	4870	AATTTCTCTGGGATCTCTCTCTGCTGATCTCATGATGATCTGTACAGCTTCACTTAA	4929
D	b	6452	AAC TTCTTAAGAAATATCATTTCTTATTAATACATGTTAAATATPATACAGTTTCAATTAA	6511
O	y	4930	TAGATCGATG 4939	
D	b	6512	AAAATGTAGG 6521	
RESULT 5				
LOCUS	A92450	4920 bp	DNA	linear PAT 22-JAN-2000
DEFINITION	Sequence 1 from Patent WO9814583.			
ACCESSION	A92450			
VERSION	A92450.1 GI:6741180			
KEYWORDS	.			
SOURCE	unidentified.			
ORGANISM	unclassified.			
REFERENCE	1 (bases 1 to 4920)			
AUTHORS	Pan,W. and Bujard,H.			
TITLE	METHOD FOR PRODUCING RECOMBINANTS INTENDED FOR USE IN A COMPLETE			
	MALARIA ANTIGENE GP190/MSP1			

JOURNAL Patent: WO 9814583-A 1 09-APR-1998;
PAN WEIQING (DE); BUJARD HERMANN (DE)
FEATURES Location/Qualifiers
source 1. 4920
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2194 a 589 c 679 g 1458 t
ORIGIN
Query Match 57.0%; Score 2814.4; DB 6; Length 4920;
Best Local Similarity 73.3%; Pred. No. 0;
Matches 3604; Conservative 0; Mismatches 1316; Indels 0; Gaps 0;
QY 10 ATGAAAATCAATTTCTCTCTGTCATTTCTGTTTTATCATCAATCTAGTGGTG 69
DB 1 ATGAAAGATCAATTTCTTTTATGTCATTTCTTTTATTAATAATACAAATGTGTA 60
QY 70 ACCCGAATCCATATCAGGAGCTGTGAAGAACTGSAAGCTTGGAAAGTCCGCTCT 129
DB 61 ACACATGAAAGTTATCAAGAGCTGTCAAAAACCTAGAACCTTTAGAGATGCAATTTG 120
QY 130 ACCGATACAGCTGTCCAGAAAGAGAGATGCTGCTGATGGAAGGAGAGTGGCAGC 189
DB 121 ACAGGTTATATGTTATTTCAAAAAGAAAAATGTTAATAATGAGAAACAAGTGGACA 180
QY 190 GCGGTTACAACACGACACCCGGTCTAAAGGCTGTGCTAGCGGTGCTCGGTGG 249
DB 181 GCTGTACACATAGTACACCTGGTCAAGGGTTCAGTTCTTCAGGTGTTAGTGGC 240
QY 250 TCTGTGCTCTGCGGGTTCCTGCTGCTCGGCGGCGGAGCTGGCATAGTGGTCA 309
DB 241 TCAGTGTCTAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 310 GCAGGCGGCGGTTCCGGGAACAGTCGAAAGAACCAATCTGACACTAGCGATTCC 369
DB 301 GCTTCAGGTGCTGAGTATTCAGAGCTGACCAATCTTCAGATTAATTTAAGGATTCA 360
QY 370 GACGCCAAGTCTGAGCCGACCTCAAGCACCAGTGAGAACTATCTCTCACTATCA 429
DB 361 GATGCTAAATCTTACGCTGATTTAAACACAGAGTACGAATTTACTTTAACTATCAA 420
QY 430 GACGTGAGTACCCACAGTGTGCAAGCTGCAATCATATGCTGACACTGCTGATAC 489
DB 421 GAACCTCAAAATTCCTCAACTCTTTGATTTAAGTATCATATTTGTTATTTGATTAAT 480
QY 490 ATTCAATGCTTCAAAATATCTGATTGACGTTAGCAAGATGATCAATGAACTCTGTAC 549
DB 481 ATTCAATGCTTCAAAATATTTAATTGATGATGATGAAATTTAATGATTTATATATA 540
QY 550 TTGAATTTCTACTTCGACTGTAGAGGCCAACTGATGACGTTTGGCCATGACTAT 609
DB 541 TTAACTTTTATTTGTTTATTTAAGAGCAAAATTAATGATGATGCTTAATGATTAAT 600
QY 610 TGTCAATTCATTCATTAATTTGAGATCAGAGCAAGCAGTGTGACGATGGAAGATTG 669
DB 601 TGTCAATTCATTCATTCATTAATTTGAGATGATGATGATGATGATGATGATGATG 660
QY 670 GTCTGCGATATGCAAGCCTCTGCAACATCAGCAATGAGCAATGTGGAAGATGGAAT 729
DB 661 GTCTGCGATATGCAAGCCTCTGCAACATCAGCAATGATGATGATGATGATGATGAT 720
QY 730 TATATTTAAAAAGATTAAGAAAGCATTGAGACATTTAAGAGCTGATGAGAAATCCAAA 789
DB 721 TATATTTAAAAAGATTAAGAAAGCATTGAGAAATTAATGAAATGAAAGATGAG 780
QY 790 AAACCATATGCAAAATTAAGATGCAACCAAGAGAGAAAAAGAAAGTGTACAG 849
DB 781 AAACCATATGCAAAATTAAGATGCAACCAAGAGAGAAAAAGAAATTAATGACAA 840
QY 850 GCCAGTACAGCTGTGCTATCTATACCAACAGCTTGAAGAGGCCCTAATCTCATGAG 909
DB 841 GCTCAATATGATCTTCTATTTTACATTAACATTTAGAGAGAGCATTAAATTAATAGC 900

QY 910 GTACTGAGAAAGCCGATAGACACCCCTCAAGAGATTAATAATATCAAGAGACTGCTGAC 969
DB 901 GTTTTAAAAAATGATATGACACTTTAAAAAATTAAGAAATTAAGATTAATGAT 960
QY 970 AAGTATATGAAATTAAGAAATCCCGCCAGCCCACTCTGGGAACACCCCTAACAGCTG 1029
DB 961 AAGTATATGAAATTAAGAAATCCCGCCAGCCCACTCTGGGAATTAACCAATATCTCTC 1020
QY 1030 CTGACAAAGAAAGAAAGATAGAGAGACAGAGAAAGATCAAAAGATGCCCAAAACC 1089
DB 1021 CTGATATAGAAACAAAAATCGAGAGACAGAAAAAGAAATTAAGAAATTCGCAAAACT 1080
QY 1090 ATTATGTTCAACATAGATCTCTCTTACTGATCCCTTGAGCTGGAGTACTCTGAGA 1149
DB 1081 ATTAAATTTAATTTGATGATTTATTTATTTATTTATTTATTTATTTATTTATTTAGA 1140
QY 1150 GAGAGAAATTAAGATATAGACATCTCCGCAAGTGCAGCAAAAGAAATCAACGAACT 1209
DB 1141 GAAAAAATTAAGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1200
QY 1210 AATGAATATCCCAATGAGTGTGAGTACCTCTGCTTATACGATATCAACAAAGCTCTC 1269
DB 1201 AATGAATATCCCAATGAGTGTGAGTACCTCTGCTTATACGATATTTAATGATTTA 1260
QY 1270 AACGACTCAATAGCTTCTGCTGATTTAACCCTTGCATTTATACGAAAGAACTCT 1329
DB 1261 AATGAATATTTCTTCTGCTGATTTAATTTATTTATTTATTTATTTATTTATTTATTT 1320
QY 1330 AAGATATCTACACAGCAATGAGAGAAAGATTTTCAACGAAATCAAGAGAAATC 1389
DB 1321 AAAAAATATATCTGATTAATGAGAGAAATTTCAATTAATGAAATTAAGAGAAATTT 1380
QY 1390 AAAATGAGAGAAAGAAATTTGAGAGTGCAGAGAAAGTTACGAAAGCCGACAGAAAGT 1449
DB 1381 AAAATGAGAGAAAGAAATTTGAGAGTGCAGAGAAAGTTACGAGAGACAGATCAAGT 1440
QY 1450 CTAAAGATATCTCTAAGAGTATGAAAGCTGCTGACAGATCTATGATTTCCAAATTC 1509
DB 1441 TTAATATATTAACAAAGAAATTTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
QY 1510 AACAATATCAATCTGACCTGACCAACTTGAGAAATGATGGAAGAAAGCTGCTTACAA 1569
DB 1501 AATATATATTTAGTTTAACTAATTTGAAAAATGATGAGATTTTCAATATATA 1560
QY 1570 GTGAGAAATCTGACACACCACTAATACCTTTGCATCTATGAGATTTCTTAACAAATCT 1629
DB 1561 GTTGAAGAAATCTGACACCACTAATACCTTTGCATCTATGAGATTTCTTAACAAATCT 1620
QY 1630 GAGAGCTCACAAAGCTTTAAGTATATGAGAGACTATTTCTGCGGAACTTTGTTGT 1689
DB 1621 GAAAGGTTAAACAAAGCTTTAATATATATGAGATTTCTTTAAGAGATTAATGATTA 1680
QY 1690 GAGAAAGAACTAAAGTTTCAAGAAATCTAATGATTAAGTAAAGCAAGAGATGAGAC 1749
DB 1681 GAAAAAGAAATTAATATTTTAAAAATTTAATAGCAAAATGAAAAATGAGATTTGAACA 1740
QY 1750 CTGTTGAGAAATTTAAGAGATGAGAAACAGTGTGTTGAGAGAAATTTAAGAAAGAC 1809
DB 1741 TTAGTTGAAATTTAAGAAATGAGAGAAACAGTGTGTTGAGAAATTTAAGAAAGAC 1800
QY 1810 GAAATTAACCAAGATGAGAGATCTGAGAGCTCCGATATTTGTTAAAGTCAAGTGCAG 1869
DB 1801 GAAATTAACCAAGATGAGAGATTTTGAAGTATCTGATGTTGAAGTCAAGTGTCA 1860
QY 1870 AAGTGTCTCTCAATGAAAGATTTGAGAGCTCAAGAGAGCTCAACGATTTGGAAGAC 1929
DB 1861 AAGTGTCTCTCAATGAAAGATTTGAGAGCTCAAGAGAGCTCAACGATTTGGAAGAC 1920
QY 1930 GTGAGTTAAACATATATATACATGTGCGCAATAGTATATAGAGAGAGATTAAGAGAA 1989
DB 1921 GTGAGTTAAACATATATATACATGTGCGCAATTTCTTACAAAGAGAAATTAAGAGAA 1980
QY 1990 CCATTAATCTCTCAATGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTA 2049

Db 1981 CCTATTATTTAAATGTGTGAAAAAAGAAATATGATTAATTAAAGTGTTCATGGCTAAG 2040
Qy 2050 GTCGAGAGCCGTGATCAGCAAGAGAGAAACATTAAACTGGAAGACAGTGCAGATAAC 2109
Db 2041 GTAGATCATTTGATTAAGAGAAAAAACAATAAAAACAGATGCAATCGGATTAAT 2100
Qy 2110 TCCGAGCCTTCCAGAGAGAGATTAACCGAGCGTACCAAGCCGAGCCGAGCAACAG 2169
Db 2101 TCGGACCATTAACCGAGAGAGAAATTAACAGCAACACTACAAACCTTGAGACACAA 2160
Qy 2170 GCCGGTTCAGCTTCGAGAGCGATAGCTGCAGAGCTCAAGCAAGAGCAGAGAGCA 2229
Db 2161 GCAGATCTGCTTTGAAGGAGATTCAGTACAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2220
Qy 2230 CAGCTCCAGTGCAGTCCGCTTCAGAGGCTAAAGCTCAAGTCCATACCCAGCT 2289
Db 2221 CAGCCAGGACGACGATCCAGTACAGAGCAAGCAAGCAAGTCCAGCCAGCAGCA 2280
Qy 2290 CCTGTGAATTAACAGAGCAGAGATGTCAAGAACTGAGTACCTTGAGAGCTCTATGAG 2349
Db 2281 CCAGTAAATTAATTAACAGTGAATATTTCCAAATTAATGATTTGAAAAATTAATGAA 2340
Qy 2350 TTCTGAATACATCCTACATCTGCCAATAATATCTCTCTCAGAGCACTATGAG 2409
Db 2341 TTTTAAATACCTCATATATATGTCAAAATATATTTGGTTTCACTCACTATGAG 2400
Qy 2410 GAGAAATTTCTTAACACTACAGATTAACCAAGAGAGAGATTAAGTCTCTTGT 2469
Db 2401 GAAAAATTTTAAACAAATTAATTAACAAAGAGAGAAAGCAATTAAGTTCAGT 2460
Qy 2470 GATCCACTGAGACCTGCTTCATATATCAAGAACATTTCCGATATGATTTCTATGTT 2529
Db 2461 GATCATTAGACTTTATTTTATATACAAATTAACATACCTGTAATGATTTCTATGTT 2520
Qy 2530 GATGACCTCAACATTTCTCTCTCACTGTTCAATGAGATATAGAGAGAGATGTC 2589
Db 2521 GATGACCTTAAACAATAGTTTATCAACATTTATGAAATTTATGAAAAAATGTT 2580
Qy 2590 TGCACCGTATTAACCAAGAGACAGCAAGAAATTAAGAACTCTGAGAGAACTAG 2649
Db 2581 TGTATTTATTAATTAAGAGATATAGCAAAATTAATTAATTAATTAATTAATTAAT 2640
Qy 2650 AAGGTCTCCACCTCTGTTAAACTCTCTTCCAGCTCCAGCAACCACTGCTCTCA 2709
Db 2641 AAAGATCCACATCTGTAAAACTCTTCAAGTATCATCAACCACTATATCATTA 2700
Qy 2710 CCTCAGACAAAGCCGAGAGAGCGCTAACGACGACCTCTCACTGCAACCTTAAT 2769
Db 2701 CCTCAGATTAACCCGAGAGTAAAGTCAATGATGATCATCTCAATTTCAAAATTTGAA 2760
Qy 2770 AACACACGAAACGTGTTGAGAACATCTGCTCTCGGAGAGAAATTAAGAACTTACCA 2829
Db 2761 AATAGTTTAATTAATTAAGAAACATATGATTTGAAAAAATTAATTAATTAATTA 2820
Qy 2830 GAACCTATTGAGAGAAATCTCCGAGAACTTCTAGGAGAGATCTGGAAGAGCGAGC 2889
Db 2821 GAATTAATAGTCAAAAAAGTAGTAAACCTTTATGAAAAATTTAAAGATAGTGT 2880
Qy 2890 ACATTTCTAAGAGAGAGCTTCACTAACTTCTGAAATCTAAAGCCGATGATCAACT 2949
Db 2881 ACATTTTATATGATCTTTTACAAATTTTGTAAATCTAAAGCTGATGATTTATTTCA 2940
Qy 2950 CTTAACGATGATCTAAAGAGAGCTGAGAGAGACATCAATTAAGCTGAAGAGCA 3009
Db 2941 TTGATGATGATCAAAAAAGAGAAATTAAGAGAGATTAATTAATTAATTAATTA 3000
Qy 3010 CTGACAGTACCTGAGCTCTTACAAAGAGTAAAGAGAACTGAGAGAGAGAGAGAG 3069
Db 3001 TTACGTTATATTTGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 3060
Qy 3070 AAGAAGAGAGAGAGAGAGAGTAAAGATGAGATCAAGAGAGAGAGAGAGAGAG 3129

Db 3061 AAAAAAAGACGTTGTAATATATAATGCAAAATTAATAAAACTTACTTATTAAGAA 3120
Qy 3130 CAGCTGGAAGAGAACTCACTGAAACAATCCGAAACAGTACTGAGAACTTCTCA 3189
Db 3121 CAATTAGATCAAAATTTGAAATCACTTAATAACCAAGAGATGATTAACAAACCTTCT 3180
Qy 3190 GTGTTCTTCAAG 3249
Db 3181 GTTTTCTTAAACAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Qy 3250 ACCAAGATTTCTTCAAACTACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3309
Db 3241 AAAAAATTTATTAAG 3300
Qy 3310 CCTGAG 3369
Db 3301 CCATTAAAACTTAAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
Qy 3370 AACTCAAG 3429
Db 3361 AACTTAAAGATTTAAGTAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Qy 3430 AAGAGCTCAGCTACCTCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3489
Db 3421 AAAAAATTTATTAAGTAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
Qy 3490 ATTAAG 3549
Db 3481 AAAAAAATTAATTAAG 3540
Qy 3550 GCAGTGAATCTTAAAG 3609
Db 3541 GCATTAGAACTTAAAG 3600
Qy 3610 GAATGAGCTCCGAG 3669
Db 3601 GAAATGAGTCCGAG 3660
Qy 3670 GAG 3729
Db 3661 GAG 3720
Qy 3730 ATCATTTGAG 3789
Db 3721 ATCATTAATTAATTAAG 3780
Qy 3790 ACCGAG 3849
Db 3781 ACAG 3840
Qy 3850 TACGAG 3909
Db 3841 TATGAG 3900
Qy 3910 GAG 3969
Db 3901 GAAATTAATTAATTAAG 3960
Qy 3970 AAG 4029
Db 3961 AAG 4020
Qy 4030 TCCCTTAACCTGAG 4089
Db 4021 TCAAGTAAATTTATGAG 4080
Qy 4090 TTTCTGATGATTAACAG 4149
Db 4081 TTTCTTAACAG 4140
Qy 4150 GATGAG 4209
Db 4141 GATGAG 4200

4210 AAAAGTATATCAAGCATAGCAAGCGAGATGAAATAATATGCGCTTCTGTAATAC 4269
4201 AAAAAATATATCAAGCAAGCAAGCGAGATGAAATAATATGCGCTTCTGTAATAC 4260
4270 ATCGAAGCCCTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4329
4261 ATTGAGACCTTATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320
4330 GCCAAGCTCTCACTATCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4389
4321 GCAAAAGTCTTAAT 4380
4390 AACTACTCAAAACCAATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4449
4381 AATTACTTAAAAACCAATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4440
4450 GGAATTCAG 4509
4441 GGAATTCAG 4500
4510 ACGGATGCTGTTCAG 4569
4501 ACGGATGCTGTTCAG 4560
4570 CTGAG 4629
4561 TTGCAAGAGATGTTAAACATTTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620
4630 ACGGCTGTTTCAAG 4689
4621 TCTGAGATGTTTCAAG 4680
4690 CAGAGAGAGAGATAG 4749
4681 CAGAGAGAGATAG 4740
4750 TGTGAGAGAGATAG 4809
4741 TGTGAGAGAGATAG 4800
4810 TGTGAGAGAGATAG 4869
4801 TGTGAGAGAGATAG 4860
4870 AATTGCTGAGAGATAG 4929
4861 AACTTCTTGAAGAT 4920

RESULT 6
LOCUS A04562 5760 bp DNA linear PAT 13-JUL-1993
DEFINITION P.falciparum P.195 gene.
ACCESSION A04562
VERSION A04562.1 GI:410754
KEYWORDS Plasmodium falciparum.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE 1 (bases 1 to 5760)
AUTHORS Holder, A.A., Sandhu, J.S., Odink, K.G., Lockyer, M.J. and
Riverso-Moreno, V.
TITLE Cloning of DNA for protozoal antigens
JOURNAL Patent: EP 0154454-A 2 11-SEP-1985;
THE WELLCOME FOUNDATION LIMITED
FEATURES
source
1..5760
Location/Qualifiers
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
gene
216..5180
/gene="P.195"
CDS
216..5180

/gene="P.195"
/codon_start=1
/protein_id="CA00367.1"
/db_xref="GI:410755"
/translation="MKIIPFLCSFLFIITNOCVTHESVOELYKLEALEADAVITCS
LFOEKEMVLNEGISTAVITTPGSGVASGSGVASGSGVASGSGVASGSGVASGSGVAS
GSGNSRTPNSDSDAKSYADLKHYRYLYLTIKELYPOLDELNHLNLTCDN
IHGFKLIDIGEINELLYKLFNEFDLRAKINDVCANDYCOIPFNLKIRAMELDYK
KLYGKRPDLKINDYKMEYIKKRNKTINELINEIDIESKRTIDKRNPNPANS
KLYOAYDLSTYNKOLEFAHNLISVLERIKIDILKRNKELKIDILKIDKIDKIDKID
NTPNLTIDKRNKIDIEBEKEIKELAKITRNKIDSLPDELELEYLBEKRNKIDISKV
ETKESTEPNEYPNGVTPPLSYNDINNALNEHLSFGLINPEITPEPSNITDNRN
KEINLEKIKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIE
PEKMGKRYSYVEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEK
YKILSKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEK
MKIDELKIKOLILKNEVLEKHNHVNPNKORNEKOPRYLLIYKIEKIDKIDKIDKID
SLINEKKNITKOGSDNESPSTEGETIQOATITKBOQAGSLLEGDSVOAOEOKOA
OPPEVPEAKAOVTPPAVNNKTEKSLDYLEKLOFLNLTSTGKTLVSHST
MNEKILKOKITKEEESKLSDDPDLKLNIONNIPVMSKFDLSNIVYHNLWYK
EIGYVEILMEIYEKEMVONLKLKDNDRILKILNEKAKVSTSVKTLSSSQPLSLT
PODRPEVSNADPESHSTLNNSLKEENILSLGKNINIOELIGOKSENFEKILK
SDPFNESTNPFVKRADINDSLNDESKRLKEEDINKLITLIDSEFLDYNKYKLE
RLPDKKTYKYMOKITLILKEOLEKSLNINPKNHYLONPSVEFNKKEAET
ENTLENTKILKHYGLVYTYNGESSPLTLSEESIQEDNVALENFVLSLEBK
KONLNEKRLSLYSLRHLHLAELEKVNKNYTGNSVNNTDYNALNEKLEFL
EOTDAVTVSESGSDTLLEQSOPEKPASTHVGAESENTITTSQWDEVDVLIYLFGE
SEEDYDLQOVVGEAVTTSVIDNLSLKEENEVLYLPLAGVSKLSEKOLENNVT
FNNVNDILNSENREKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEK
VNYIKDSIDTDLNFAVDYGYKILSEKYSKDSLSIKYIKNDKOGENEKYLPLNIE
TYKTYNOKIDILFVHLEKAVLYNTEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEK
GIADLSTOYNHNNLILKFLSTGVNPEFNILKSVSLNLDKRIARYVHHTPPARKNMI
QOSSGCFRNLDEREKECKLNLKROBESKVEKNSPNCNNGGCDADACTEDSGSN
GKRTCECKPPDYPPLSMVTFSSSNFLDISFLILMLILYSFI"

BASE COUNT 2565 a 730 c 725 g 1840 t
ORIGIN

Query Match 54.8%; Score 2706.2; DB 6; Length 5760;
Best Local Similarity 72.1%; Pred. No. 0;
Matches 3587; Conservative 0; Mismatches 1343; Indels 45; Gaps 3;

10 ATGAAGTAT 69
216 ATGAAGTAT 275
70 ACCGACGAATCTATAG 129
276 ACACATGAAGATATCAAGAACTGTCAGAACTTGAAGAGATGAGATATG 335
130 ACCGATACAGCCCTGTCAG 189
336 ACAGGTTATAGTTATATCAAG 395
190 GCGGTACACACAGACACCCGCTTCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 249
396 GCTGTTCAACATGATACCTGTTTCAAG 455
250 TCTGTGCTCTGAG 309
456 TCAAGTCTTCAAG 515
310 GCAAGCGGCTTCCGAG 369
516 GCTTCAG 575
370 GAGCCAG 429
576 GATGCTAAATCTTACGCTGTTTAAACACAGAGATAGAGAAATTAATCTGTTATACATATAA 635
430 GAGCTGAAGTACACAG 489
636 GAACCTCAATATCTCAACCTTTGATTAATCAATATATATATATATATATATATATATATATAT 695
490 ATTGATGCTTCAATATCTGATTTGACGTTAGAGAGAGATCAATGACATCTCTGTACAG 549


```

Db 1776 GTTGAGAACTTACACACCCTTAATCTTGGCATCTATGAAAAATTCATAATCTT 1835
Qy 1630 GAGAGCTCACCAAGCTCTTAATATATGAGAGACTATCTCGCGGAACATGTGTG 1689
Db 1836 GAAAAATTAAACAAAGCTCTTAATATATGAGAGACTATCTTAAAGAAATAGTAGT 1895
Qy 1690 GAGAAAGAACTAAAGTATTTACAAAGATCTCATAGTAAAGTGAAGAGAGATGAGACG 1749
Db 1896 GAAAAAGATTAATATATTAATTAATTAAGCAAAATAGAAATAGATTTGAACA 1955
Qy 1750 CTGTGAGAACATTAAGAGATGAAGAACAGTTGTTGGAAGAAAGATTACAAAAGAC 1809
Db 1956 TTAGTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2015
Qy 1810 GAAATTAACCAAGATGAAGAGATTCCTGAGAGTCCCATATCTTAAGTCAAGTCCAG 1869
Db 2016 GAAATTAACCAAGATGAAGAGATTCCTGAGAGTCCCATATCTTAAGTCAAGTCCAG 2075
Qy 1870 AAGTGCTCCATGAACAGATTAAGTGAAGTCAAGAGTCAAGTCAAGTCAAGTCAAG 1929
Db 2076 AAAGTTTATTAATGAACAAATTAAGAGATTAATTAATTAATTAATTAATTAAT 2135
Qy 1930 GTGAGTTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1989
Db 2136 GTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2195
Qy 1990 CCATACATACCTCATCTGATCTCAAGAAAGATTAACAACTGAAGTTCATGCCCAA 2049
Db 2196 CCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2255
Qy 2050 GTGAGAGCTTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2109
Db 2256 GTAGAAATCATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2315
Qy 2110 TCGAGCTTCACAGAGAGAGAGAGATTAACGAGAGCTTACACCAAGCCCGCAACAG 2169
Db 2316 TCGGAACCATCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2375
Qy 2170 GCGGATTCAGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2229
Db 2376 GCAGAGATCTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2435
Qy 2230 CAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2289
Db 2436 CAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2495
Qy 2290 CTTGTGAATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2349
Db 2496 CCAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2555
Qy 2350 TTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2409
Db 2556 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2615
Qy 2410 GAGAGAGATTCCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2469
Db 2616 GAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2675
Qy 2470 GATCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2529
Db 2676 GATCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2735
Qy 2530 GATGAGCTCAACA-----ATTC 2547
Db 2736 GATGAGCTTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2795
Qy 2548 CTCTCTCACTGTTCAATGAGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 2607
Db 2796 GTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2855
Qy 2608 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2667
Db 2856 AAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2915

```


Db 5076 CCTGATTTGTAATCCACTTCATGAGTAAATTTCTCAGTTCCTTAACATTTTACGAATA 5135
Qy 4885 TCCCTCCGCGATGCCATGCTGATCCTGATACACTTCATCTAATGATGATG 4939
Db 5136 TCATTCTTATTAATACATCATGTTAATATTAACAGTTTCATTTAAAAATGTAGG 5190

RESULT 7
E00656
LOCUS E00656 5760 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding p.195 protein of plasmodium falciparum.
ACCESSION E00656
VERSION E00656.1 GI:2168935
KEYWORDS JP 1986019490-A/1.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 5760)
AUTHORS Ansonli,A.H., Maikuru,J.R., Jiyasubida,S.S., Barentelina,R.M. and Karenu,G.O.
TITLE DNA CLONING FOR ANTIGEN OF PROTOZOA
JOURNAL Patent: JP 1986019490-A 1 28-JAN-1986;
COMMENT WELLCOME FOUND LTD:THE
OS Plasmodium falciparum
PN JP 1986019490-A/1
PD 28-JAN-1986
PR 22-FEB-1985 JP 1985034280
PR 22-FEB-1984 GB 84 8404692, 26-SEP-1984 GB 84 8424340 PI
ANSONLI AASBA HORDUPA, MAIKURU JIEMUSU ROTSUUIYAA, PI
JIYASUBINA SHINKU SANDEYU, BARENTEINA RIBEROSU MORENO, PI KARENU
GERITSUTO ODEINKU
PC C12N15/00,A61K39/015,C07H21/04,C07K13/00,C12N1/20,C12P21/02,
PC C12N1/20,
PC C12N1:19),(C12P21/02,C12R1:19);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH key Location/Qualifiers
FT CDS 216..5180
FT sig_peptide 219..272
FT mat_peptide 273..5177
FT Location/Qualifiers
1..5760
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
BASE COUNT 2565 a 630 c 725 g 1840 t
ORIGIN

Query Match 54.8%; Score 2706.2; DB 6; Length 5760;
Best Local Similarity 72.1%; Pred. No. 0;
Matches 3587; Conservative 0; Mismatches 1343; Indels 45; Gaps 3;

Qy 10 ATGAATATCATTTCTCTCTGTCATCTGTTTATCATCAATATACAGTCGTG 69
Db 216 ATGAAGATCATATTTCTTTTATGTTTATCTTTTATATTAATAACAAAGTCTA 275
Qy 70 ACCCAGCAATCTTATCAGAGCTGTTAAGAACTGAAGCTTGGAGATGCCCTT 129
Db 276 ACACATGAAGATTTACAGAGCTGTCAAAAACCTAGAGCTTTAGAGATGCAATATG 335
Qy 130 ACCGATACAGCTGTTCCAGAGAGAGATGCTGTAATGAAGAGACAGATGCCAG 189
Db 336 ACAGATTATGTTATTTCAAAAGGAAAAAATGATTAAATGAAGCAAAAGTGAACA 395
Qy 190 GCCGTTAACACAGACACCGGTTCTAAAGGCTGTGAGGCTAGCGGTCCCGTGG 249
Db 396 GCTGTACACTAGTACACCTGGTTCACAAAGGTTCAAGTGTTCAGGTTCAGGTGC 455

Qy 250 TCTGTGCCCTCGGGGGTTCCTCCCTCCGGCGGACGCTGGCATCAGTGGCTCAGT 309
Db 456 TCAATGCTTCAGTGCGCTCACTTCTCAGTGGCTCAGTGGCTTCAGTGGCTCAGT 515
Qy 310 GCAGCGGCGGTTCCGGGAACAGTCGAAGAACCAATCCATCTGACACTAGCGATTC 369
Db 516 GCTTCAGGTGGTTCAAGATTAATTCAGAGCTACAAATCTTCAGATTAATTCAGATTC 575
Qy 370 GACGCCAAGTCTACGCCGACCTCAAGACCGAGAGAAATATCTCTCTATCAAG 429
Db 576 GATGCTTAATCTTACGCGATTTAAACACAGATGACAAATTAATCTTAACTTAACT 635
Qy 430 GAGCTGAAGTACCCACAGTGTGACCTCACTATATCATATCTGACACTGTGATAC 489
Db 636 GACTCAATATATCTCAACTCTTTATTTAACTATATCTATCTTGTGTATAT 695
Qy 490 ATTATGCTTCATATATCTGATTTAGCGTTACGAAGATCAATGAACTCCTGTACAG 549
Db 696 ATTATGCTTCATATATTTAATTTATGATGATGATGAAGAAATTAATGATTAATAA 755
Qy 550 TTGAATTTCTACTGACTTGTCTAAGGCGCAACTGAATGACGTTTGGCCCATGAT 609
Db 756 TTTAACTTTTATTTGATTTATTAAGACCAAAATTAATGATGATGCTAATGATTA 815
Qy 610 TGTCAATTTCCATCAATTTTGAATGATGACGACCAAGCTTGTGAGTGAAGAGTTG 669
Db 816 TGTCAATTTCCATTTCAATTTTGAATTTGATGATGATGATGATGATGATGATGAT 875
Qy 670 GTCTTCGATATCGCAAGCTCTCGACCAATCAAGCAATGTGGAGAGATGAGAT 729
Db 876 GTCTTCGATATGAGAAACCATATGACATTAATTAAGATTAATGTGAGAAATGAGAT 935
Qy 730 TATATTAAGATTAAGAACCCATGAGAACATTAACGAGCTATGAGATTCAAA 789
Db 936 TACATTAAGATTAAGAACCCATGAGAACATTAATTAAGATTAATTAAGATTAAG 995
Qy 790 AAGACCATAGACAAATTAAGATTAAGAACCAAGAGAGAGAGAGAGAGAGAGAGAG 849
Db 996 AAACCAATTTGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 1055
Qy 850 GCCCATGACAGCTGTGCTATTAACAAACAGCTTGAAGAGCCCATTAACCTATCAGC 909
Db 1056 GCTCAATTTGATTTCTTCTATTTACAAATTAAGATTAAGAGAGAGAGAGAGAG 1115
Qy 910 GTACGAGAGAGCGCATGACACCTCAAGAAATTAAGATTAAGATTAAGATTAAG 969
Db 1116 GTTTAGAGAGAGCGATTTGACCTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAT 1175
Qy 970 AAGATTAATGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1029
Db 1176 AAGATTAATGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1235
Qy 1030 CTGACAG 1089
Db 1236 CTGATTAAG 1295
Qy 1090 AATTAATGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1149
Db 1296 AATTAATGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1355
Qy 1150 GAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1209
Db 1356 GAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1415
Qy 1210 AATGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1269
Db 1416 AATGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1475
Qy 1270 AACGAGCTTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1329
Db 1476 AATGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1535
Qy 1330 AAGATTAATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 1389

Db	1536	AAAAATATATCTACTGATTAATGAAAGAAAAAAATTTATATAATGAAATTTAAGGAAAAAATTT	15935
Qy	1390	AAAAATGGAAGAGAAAAATTTGAGCTGACAAAGAAAAGTTACGAAGACCGCAGCAAAAGT	14449
Db	1596	AAAAATGAAAAAATAAAATTTGAATCTGATTAATAAAATCTTACGAAGACAGATCTAAGTCT	16555
Qy	1430	CTAAACCATATCACTAAAGAGTATGAAAAGCTGCTGACAGACATCTATGATTTCCAAATTC	15099
Db	1656	TTAAATGATTAATACAAAAAGAAATATGAAAAATTAATTAAAGAAATTTATGATAGCAAAATTC	17151
Qy	1510	AACATATAACATCGACCTGACCACTTCGAGAAATATGCGAAAAACGGTACTCTTCACNA	15669
Db	1716	AATATATATATAGTTTAACTAATTTTCGAAAAATGATGGTAAAGAAATTCATATATAA	17755
Qy	1570	GTGAGAAACTGACACACCATTAATACCTTTGCATCTATAGAAATTTCTAAGCATATCTT	16299
Db	1776	GTGAGAAACTTACACACCTTAATACCTTTGCTATCTGCTAAGAAATTTCTAAGCATATCTT	18355
Qy	1630	GAGAAAGCTCCCAAGCTCTTAAGTATATGAGAGCTATATTCCTCGAGACCTTGTTCG	16899
Db	1836	GAAAGATTAAACAAAAGCTCTTAAATATATGGAAGATTATCTTTAAGAAATTAATATATTT	18955
Qy	1690	GAGAAAGACTAAAGTATTAACAAGATCTCATAGTAAGATGGAAGACGATCGACAGC	17449
Db	1896	GAAAGAAATTTAAATATTTTATAAAATTTTATAAGCAAAATAGAAATAGAGATTGAACCA	19555
Qy	1750	CTTGTTGAGAACTTTAAGAGATGGAAGAACAGTTGTTTGAAGAGACATTTACAAAGAC	18099
Db	1956	TTAGTTGAAATATTTAAAAAGATGAGAAACAGCTTTTGAAAAAAAATTTACTAAAGAC	20155
Qy	1810	GAAATTAACAGATGGAAGATCTCGAGAGCTCTCCGATATTGTTAAGTCCAGCTGAG	18699
Db	2016	GAAATTAACAGATGGAAGATTTTGAAGATCTCGAGATTTGTTAAGTCCAGCTGAG	20705
Qy	1870	AAGTGCTCCCTCAGACAGATTTGATGAACTCAACAACTCACTCATCTGTAAGAAC	19299
Db	2076	AAAGTTTTATTAATGAACAAATTTGACGATTTAAAAAGACTCAATTTGATTTTAAAAAAT	21355
Qy	1930	GTGAGATTTAAAACATTAATATACATGTCGCGAATAGTTATAGCAGAGATTAAGCAGAA	19899
Db	2136	GTAAGATTTAAAACATTAATATACATGTTCCCAATTTCTTCAACAAAGAAATTAAGCAGAA	21955
Qy	1990	CCATATCACTCACTGATCACTCAAGAAAAGATATGACAACTGAAAGTGTCAATGCCCAA	20499
Db	2196	CTTATTAATTTAATTTGTTGAAAAAAGAAATTGATTAATTTAAAGTGTCTATGCTAAG	22555
Qy	2050	GTCAGAGCCGTGATCAAGAGAGAAAGAAACATTAACATTAACATGAAGAGACATCATATAC	21099
Db	2256	GTAAGATCATTTGATTAATGAGAAAAAATAAACATTAACAAACCAAGTCAATCGATATAT	23155
Qy	2110	TCCGAGCCTTCACAGAGAGAGATTAACCGGACAGGCTACCAACCAAGCCCGGATCAACAG	21699
Db	2316	TCCGAGACCATCAACCGAAGAGAAATTAACAGGACAGCACTATCAAAAACCTGGAGACAA	23755
Qy	2170	GCCGCTTACGCTCTCAAGGCGATACGCTGCAAGCTCAAGCAACAAAGACAGACAGCACA	22299
Db	2376	GCAGGATCTGCTTTAGAGAGAGATTCATGATCAACAGCAACAAAGCAACAAACAAACACA	24355
Qy	2230	CAGCTCCAGTCCGAGGCGCCGTTCCAGAGGCTTAAGCTCAAGTGCCTACACCCACACT	22899
Db	2436	CAACCCACGATCACGATCAACGATCAAGAGCAAAAGCAACAAAGTCCCAACACACACACA	24955
Qy	2290	CTTGATTAACAAGACCGAAGATGTCAGCAAACTGATCTTACCTTGAGAAAGCTCTATAG	23499
Db	2496	CCAGTAAATTAATAACTGAAAATGTTTCCAAATTTGATTAATCTTGAAAATTAATATACA	25555
Qy	2350	TTCTTGATATACATCTTACATCTGCCCAATAATATCTCTGCTCCACAGCACTATGAC	24099
Db	2556	TTTTTAATTAATCTTATATATATGTCACAAATATATTTTGTTCACACCTCAACATATGAC	26155
Qy	2410	GAGAGATTTCTTAACACGTCACAGATTAACCAAGAGAGAGAGATTAACCTGCTCTTGT	24699

Db	2616	GAAGAAGATTTAAACAACATTTAAATTTACAAAGAGAGAAAGAAACAAATTAGATTCAGCT	2675
QY	2470	GATCCACTGCGACCTGCTGTTTCATATTCGAGACACATTCGCCGTTATGTATTCTATGTC	2529
Db	2676	GATCCACTAGACCTTATGTTTAAATATACAAAAATACATCCTGTAATGTATCTATGTTT	2735
QY	2530	GATGAGCCCAACA-----ATTCT	2547
Db	2736	GATACCTTAAACATAGTTTATCACACATATTTATGGGTTTATGAAAAAGAAATTTGGTAT	2785
QY	2548	CTCTCTCAACTGTTCCATGGAGATATATGAGAAAGAGATGGTCGACACCTGTATTAACCTC	2607
Db	2796	GATATTATATTACTTATATGAAATTTATGAAAAAGAAATGGTTTGTATTTATATTAATTAAGCTT	2855
QY	2608	AAAGACAACGACGAATTTAGAAACCTTCTGGAGAAAGCTTAAGAGSTCTCCACCTCTGT	2667
Db	2856	AAGGATATATGCAAAATTTAAATTTATAGAGAAAGCAAAAAAGTATCCACATCTGTA	2915
QY	2668	AAAACCTCTCTTCACCTCCATGCAACCCCTCTCTCAACCTCAACGACAAAGCCGCA	2727
Db	2916	AAAACCTCTTCAAGTATATCAATATGACCACTTATCTTAAACCTCAGGATTAACCCGAA	2975
QY	2728	GTGAGCCGTAAAGCAGCAGACACCTCTCACTCGACCCAACTTAATCTCACTGAAACCTGTTT	2787
Db	2976	GTAAGTGCAAAATGATGATACATCACTATTCACAAATTTGAATATATAGTTTAAATTAATTTT	3035
QY	2788	GAGAACATCTCTGTCTCTCGCAGAAATAGACATCTACCAAGAACTTATTTGGACAGAA	2847
Db	3036	GAAGAACATATTTGACTCTTGGAAAAACMAAAATATATACCAAGAAATTAATAGCTCAAAAA	3095
QY	2848	TCTGTCGGAACCTCTACGGAAGATATCTAAAGAACAGGACACATTCATATACGAGAGC	2907
Db	3096	AGTACTGAAACCTTTTATGAAAAAGATTTTAAAGATAGTATACATTTTATATGAATCT	3155
QY	2908	TTCACTAACTCTGTAATTTAAAGCCGATATCACTCTTAAAGATGAATCTTAA	2967
Db	3156	TTTACAATTTTGTAAATATCTAAAGCGATGATATTAATTCATGTAAGATGAATCAATCAAA	3215
QY	2968	CGTAAAGACTGGAAGAGGACATCAATAGCTGAGAGACACTGCACACTGACTTGAC	3027
Db	3216	AGGAAGAAATATAGAGAGAGATTAATTAATTTAAAAAACTTTACAGTTATCACTTGTAT	3275
QY	3028	CTGTAAACAAGTCAACATCGAAACGAGAGAGCTCTTGACAAAGAGAAAGCACTGGC	3087
Db	3276	TTATATATTAATATTTAATTTAAAAATTAGAAAGTTTATTTGATTAAGAAAGAAACACTTGGT	3335
QY	3088	AAATTAAGATCGATCAAGAGAGTTGACTCTGCTCAAGAGCAGCTTGAAGAACACTC	3147
Db	3336	AAATATTAATGAATTAATAAAAAACTTACTTTATTAAGAAACAAATTTGAATCAAAAAATG	3395
QY	3148	AACTCACTGAACAATCCGAAACAGCTGACAGACTCTCTCAAGTGTCTTCAACAAGANG	3207
Db	3396	AATTCACCTTATATACCAAGAGCATATGATTAACAACAACTTTCTGTTTCTTTTACAAAAAA	3455
QY	3208	AAGGAAGCCGAGATCGCCGAGACAGAGAACACTCGGAGAAACACAAAGATTTCTCTCAA	3267
Db	3456	AAAGAGCTGAAATAGGAGAACTGAAACACATTTAGAAACACAAAAATATTATTGAA	3515
QY	3268	CACATCAAAAGCCCTCGTCAGTATTTAATATGGGAGCTCTTCTCACTGAAGCTCTGCC	3327
Db	3516	CATTATTAAGACACTTGTTTAATATTTATATGATGGATCATCTTCCATTAATAAACTTTAAGT	3575
QY	3328	GAGGAGAGACTCGACGACCGAGGATTAATCTACGCCAGCCTCGAGAACCTTCAAGGCTCTGTCT	3387
Db	3576	GAGATATCAATTAACAACAGAGATATTTATGCCAGTTTGAAACCTTTAAGTATTAAGT	3635
QY	3388	AACTCGAAGGCAAGCTGAGAGCAACCTGAACTGAGAGAGAGAGAGCTCAGCTACTC	3447
Db	3636	AAATTAAGAGAAATTAATTAAGATATCTTAATTTAGAAAAAATAATTTATCATACTTA	3695
QY	3448	TCTATCGGAGCTGATCCCTGATATCGCCGAGCTCAAGAGATGATTAAGAGCAAGAACTAC	3507
Db	3696	TCAAGAGGTTTACTCATCTTTAATTTGCTGATTTAAAGAAAGTATTAATAAATAATTAATTTAT	3755

OY	3508	ACGGCAATAGGCCAACCGAGTAATPACAGACGTGAATAAGCACTGGAAATCTTACAG	3567
Db	3736	ACAGGTATTTCTCCAAACCGTAAATAATACGGATGTTTAAACAATGCATTAGAACTTTACAA	3815
OY	3568	AAGTTCCTGCCTGAAGAGACAGATGTGCGCAGCTGTGGTCTGTGAATCTGGCTCGACACA	3627
Db	3816	AAATTTCTCCAGAAAGAGACAGATGTGGCAAGCTTTGTAAAGTGAAGTGATCCGACACA	3875
OY	3628	CTGAGACAGTCTCACTTAAGAAAGCTGCATCTACTCATGTGGAGCCGATGCCAATPACA	3687
Db	3876	TTTAAACAAGAGTCAACCAAGAAAGAACAGAGATCAACTCAATGTAGGAGACAGAGCTTAACACA	3935
OY	3688	ATTACCAATCTCGAAGCTGCAGAGTGAAGTGCAGATGACGTATCATTTGTGGCTATCTTC	3747
Db	3936	ATTAAACAATCACAAATATGTCGATGATGAAGTGAAGATGACGTAAATATATAGTACTCATATT	3995
OY	3748	GGCAGAGACCGAGAGAGCTACATGACCTTCGCGCAGGTGTCACCGGTGAGGGCTGTCACT	3807
Db	3996	GGAGAAATCCGAAAGAAAGTTATGTATGATTTTAAAGCAAGTATGTAACGAGGAAGACAGTACT	4055
OY	3808	CTTCCCGGATTTGATTAATCTTGTGCCAAATGCGAACAAGATPACGAAGTGGCTATATG	3867
Db	4056	ACTTCCGTAATGTATTAACATCTTCTTAAATTTGAAATATGAATATGAGGTTTATATATTA	4115
OY	3868	AAACCTTGCGAGCGCTATAGTCTCTCAAGAAAACAGCTGGAGATTAACGTGATGACC	3927
Db	4116	AAACCTTTAGAGGTGTATTTAGAACTTTAAAAAAACAATTTGAAATAATACGTTATGACA	4175
OY	3928	TTCAATTCGAACGGAAGAGACATTTGGAACAGCGCTTTAATAAGAGAAAATTTTCAG	3987
Db	4176	TTTATGTATTAATGTAAATGATTTTAAATTCACGATTTTAAATTAACGTAATTTTCAA	4235
OY	3988	AACTCTCTGGAGAGCGAGTATTCCTTATTAAGAAAGCTGACCTCTCACTACAGCTGTG	4047
Db	4236	AAATGTTTAAGATCAAGATTTAATTCATATPAAAGATTTAACTCAAGTATATATGTGTG	4295
OY	4048	AAGAACCCATCAAGTTCCATATAAGAGAGAGGATAAATTTCTGTATAGTTACAC	4107
Db	4296	AAAGATCCATATAAATTTCTTAATAAGAAAAAAGATAAATTTCTTAAGCGTTATAT	4355
OY	4108	TATATTCAGAGCTGCATCGACACCGATATTCATTTGCTAAATGATGTGCTGGAGTTATAC	4167
Db	4356	TATATTAAGGATTCATATAGTACGGATATTAATTTTGCAAATGATGTTCTTGATATATT	4415
OY	4168	AAGATCTGACGGAATAATACAAAGTGTGACTGTATTTAAAAAGTATATCAAGAT	4227
Db	4416	AAATATTTATCCGAAATATTAATTAATCAGATTTTATGATTCATTTAAAAATATATCAAGAC	4475
OY	4228	AAGGAAGCGGAATGAAAAATATCTGCCCTTCGTAATATAATGGAACACCTGTACAG	4287
Db	4476	AAACAAGGTGAATAATGGAATACCTTCCTTTTAAACAATATTTGAGACTTATATATAA	4535
OY	4288	ACAGTGAACGCAAAAATCGACTCTTCGTATATCACTGTGAGGCCAAGGTCTCTCAACTAT	4347
Db	4536	ACAGTTATATGATTAATAATTTGATTTATTTGTATTCATTTAAGAGCAAAAGTTCTAAATATT	4595
OY	4348	ACTTACGAGAAAGACATGTGGAATTTAAAAATCAGGAGCTAATCTCTCAAAACAATC	4407
Db	4596	ACATATATGAAGATTAACAAGTAAAGTTAAATTTAAAGAACTTAATTTCTTAAAAACAAT	4655
OY	4408	CAAGACAAGCTGGAGATTTTCAGAAAAAATTAACAATTTGTGCGGAATTTGCAGACTGTCT	4467
Db	4656	CAAGACAATTTGGCAGATTTTAAAAAAAATTAACAATTTGTTGGGAATTTGCTGATTTATCA	4715
OY	4468	ACCGATATATAACAACAATCTCTGACCAAGTTTCTGTCACTGGCAATGGTGTTCGAA	4527
Db	4716	ACAAATTTATTAACCTATATTAATCTATTGACAAGATGCTTATGACAGTATGTTTGTAA	4775
OY	4528	AACTCCGCAAAAACAGTGTGAGACATCTGTCTGA-CGGCAACCTGCAAGGAGTGTCTGAA	4586
Db	4776	AACTGTCTAAATCCGTTTATCTATTAATTAATCTGATTTGGAACCTTGCAAGGATGTCTGAA	4835

4587	CATTCGCCAGCAGCAATCGTGGAAACACAGG--CCCCCAAAATGAGCGGTGTTACAG	4644
4836	CATTTCACAACACCAATGCGTAAAAAACAATGATCCACAGAGTTCTCGGATGTTTCAGA	4895
4645	CATTCGAGCAGCGCGGAAGAGTCAAGTGCTCTCTGAACCTACAAACAGAAAGAGATAG	4704
4896	CATTAGATGAAAGAGAAAGATGTAAATGTTTAAATTCACAAACAGAAAGATAGTAA	4955
4705	TGCGGAGGAACCCAAACCCATCCGCAATGAAACAAATGAGCGGTGAGCGCGATGCT	4764
4936	TGTGTTGAAATATCAATCTCACTGTGTAACGAAATATATGATGATGATGACGATGCC	5015
4765	AAATCACCAGGAGAGACAGCGGCTCTTAACGGAAAGAAATACATGCGAGTGTACTAAG	4824
5016	AAATGTACCGAAGAAGATTCAGTATGACGACGAAAGAAATACATGTGATGTACTTAA	5075
4825	CCCGACTCCATCCACTCTTGACAGCGGATTTTTCCTCCACGCTTAATTTCTGGGCATC	4884
5076	CTGTATTTATCCCACTTTCGATGGTAAATTTCTGCGACCTTCCTTAACCTTTAGGAATA	5135
4885	TCCCTTCGCTGATCCATGCTGATGCTGATCCCTGACACCTCATCTAATAGATGATG	4939
5136	TCATTCTTATTAATCTCATGTTATATATATACAGATTTCATTTAATAAATGTAGG	5190
RESULT 8		
PEP190	5282 bp	DNA linear
LOCUS	P.falciparum gpi190 (MSA1, MSP1, PMMSA)	gene for merozoite surface antigens precursor.
ACCESSION	X03371	
VERSION	X03371.1	GI:929797
KEYWORDS	gpi190 gene; merozoite surface antigens; MSA1 gene; MSP1 gene; PMMSA gene.	
SOURCE	Plasmodium falciparum.	
ORGANISM	Plasmodium falciparum	
REFERENCE	1 (bases 1 to 5282)	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	Mackay, M., Goman, M., Bone, N., Hyde, J.E., Scalfe, J., Certa, U., Stuenkelberg, H., and Bujard, H.	
TITLE	Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level	
JOURNAL	EMBO J. 4 (13B), 3823-3829 (1985)	
MEDLINE	86136024	
PUBMED	3004972	
REFERENCE	2 (bases 1 to 5282)	
AUTHORS	Pan, W., Tolle, R., and Bujard, H.	
TITLE	A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gpi190/MSA1	
JOURNAL	Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)	
MEDLINE	96123395	
PUBMED	8577332	
REFERENCE	3 (bases 1 to 5282)	
AUTHORS	Bujard, H.	
TITLE	Direct Submission	
JOURNAL	Submitted (23-JUN-1995) H. Bujard, Zentrum fuer Molekulare Biologie der Universitaet (ZMBH), Im Neuenheimer Feld 282, Heidelberg 69120, FRG	
COMMENT	On Aug 5, 1995 this sequence version replaced gi:9920.	
FEATURES	Location/Qualifiers	
source	1..5282	
	/organism="Plasmodium falciparum"	
	/isolate="K1"	
	/db_xref="taxon:5833"	
	/chromosome="9"	
	/dev_stage="blood stage"	
	72..4964	
gene	/gene="gpi190, MSA1, MSP1, PMMSA"	
	72..4964	
gene	/gene="gpi190, MSA1, MSP1, PMMSA"	
	/codon_start=1	
CDS	/product="Precursor for the major merozoite surface antigens"	

OY	1570	GTGAGAACTGACACACCCCTAATACCTTTGCATCTCTATGACAATTCCTACGATPATCTT	18239
Db	1605	GTGAGAACTTACACACCCATTAATCTTTGGCATCTCTAATGAAAATTTCTTAACATPATCTT	16644
OY	1630	GAGAACTCACCAAGCTCTTAAGTATATGAGAGCATATCTCTCCGAGCAATGTTGTG	18689
Db	1665	GAAAAGTTAACAAAAGCTCTTAAATATATGGAAGATTAATCTTTAAGGATATATCTACTT	17244
OY	1690	GAGAAAGAACTAAAGATTTTCAAGAAATCTCATAGTAAGTGAAGTGAAGACGATCGACAGC	1749
Db	1725	GAAAAGAAATTAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	17844
OY	1750	CTTGTTGAGAACTTAAAGAGATGAAAGACAGTCTTTTGACAGAGATTTCAAAAGAC	18099
Db	1785	TTAATGTAATATATTAAGAAAGATGAAAGACAGCTTTTGGAAAAAAAATTAATCTAAAGAC	18444
OY	1810	GAAATTAACACAGTGTGAAGATCTCGAGAGCTCTCGATATTTGTAAAGTCCAAAGTGCAG	18699
Db	1845	GAAATTAACACAGTGTGAAGAAATTTTATGAGATATCTGACATTTGTAAAGTATCAAGTCTAA	19044
OY	1870	AAGGTGCTCCTCATGAACAGAAATTTGATGAACTCCAGAAAGACTCAACTCATCTGAAAGAC	19299
Db	1905	AAAGTTTATTAATGAACAAATTTGACGAATTAAGAAAGACATCAATGATTTTAAAAAT	19644
OY	1930	GTGAGTTAAAACTATATATACATGTGCGCAATAGTTATTAACAGAGATATACGAGAA	19899
Db	1965	GTAAATTTAAAACTAATATATACATGTGTTCCCAATTTCTTCAAAAAGAAATTAAGCAAGAA	20244
OY	1990	CCATATACCGATACGATACACAAAGAAAGATGACAAACCTGAAGTGCATAGTCCCAAA	20499
Db	2025	CTTATATTTAATGTGTTGAAAAGAAATTTGATTAATTAAGTGTCTATGCTATAG	20844
OY	2050	GTCGAGAGCCTGATCAACGAGAGAGAAAGAAATTAACATTAACGAGACAGTCAAGATTAAC	21099
Db	2085	GTAGAAATCATTTGATTAATTAAGAAAGAAAAAAACATTAACAGAGAGTCAATCGATATAT	21444
OY	2110	TTCGAGCCTTCCACAGAGAGAGATTAACCGGACAGCTTACCAACCAAGCCGGACACAGC	21699
Db	2145	TCGGAAACCATCAACCGAGAGGAAATTAACAGGACCAAGCAACTTACAAAACCTGGACACAA	22044
OY	2170	GCGCGTTACGCTCCGAAGGCGATACGAGCAAGCTCAAGACCAAGACAGAGAGAGGACCA	22299
Db	2205	GCAGGATCTGCTTTAGAGGAGATTTAGATTAACAAGCAACAAAGCAAAAACAAAACAAACA	22644
OY	2230	CAGCCTCCAGTGCAGAGTGCCCGTTTCCAGAGGCTAAAGCTCAAGTGCCTTACACACACACT	22899
Db	2265	CAACACACAGTACAGTACCAAGTACCAAGAAAGCAAAAGCAAGCAAGTCCCAACACACACACA	23244
OY	2290	CCTGTGAATTAACAGACCCGGAATGTCCAGCAAACTGACTACTTGTAGAGAGCTATAGC	23499
Db	2325	CCATTAATATATAAACCTGAAAATGTTCACAAATTAAGATATCTTGAAAAATTAATATGAA	23844
OY	2350	TTCTCGAATACATCTCAATCTCCGCAAAATATATCTCTGCTCCACAGCACTATATGAC	24099
Db	2385	TTTTTAATATCTTCATATATATATATCTCAAAATATATTTTGTTTCACTCAACTATGAC	24444
OY	2410	GAGAGATTTCTTAACAGTACAGATTAACCAAGAGAGAGAGATTAACCTGCTCTTGT	24699
Db	2445	GAAAAGATTTAAAAACAATTAATAAATTTTACAAAGAGAGAGAAAGCAAAATTAAGTTCACT	25044
OY	2470	GATCACTGAGCCTGCTTCAATATCCAGAACACATTTCCGTTATGTATTTCTATGTTTC	25299
Db	2505	GATCCATTAAGACTTATGTTTATATATACAAAATATACATATCTGTAAATGATTTATGTTT	25644
OY	2530	GATGAGCTCAACAAATTTCTCTCTCAACGTTCATGAGATATATGAGAGAGAGATGCTC	25899
Db	2565	GATGAGCTTAACAACTACTTTTATACAACTATTTATGGAATTTTATGAAGAAAGAAATGCTT	26244
OY	2590	TGCACACGTATTAACCAAGCAAGCAAGCAAGATTAACACCTCTGAGAGAGAGCTAG	26499
Db	2625	TGTATATTTATATATACTTAAGGATATATGACAAATTAATTAATTAATTTTATGAGAGAGCGAAA	26844
OY	2650	AAGGTCTCACACTCTGTATTAACCTCTCTCTCCAGCTCCATCAACCAACTGCTCTCA	27099

Db	2665	AAAGATATCACATCTGTAATAAAACCTCTTCAAGTTCAATCATATCAACCATTAATCACTTAACA	2744
Qy	2710	CCTCAAGACCAAGCCCGAAGTAGAGGGCTAACGACGACACCTCTCACTCGACCAACTTAT	2769
Db	2745	CCTCAGGATTAACCCCGAAGTAGAGGCAAAATGATGATACATCTTCACAAATTTGAT	2804
Qy	2770	AACCTCAGAAACGTGTTTGAGAACATCTCGTCTCTCGCGAAGAAATAGAACAATCTACCA	2829
Db	2805	AATAGTTTAATTAATTTTGAAAACATATGAGCTCTGGAAAAACAAAATATATATACCA	2864
Qy	2830	GAACTTATGACAGAAATGCTCCGAGACTTACGAGAAATATCTGAAAGACGCGAC	2889
Db	2865	GAAATTAATAGGTCAAAAAAGTAGTGAACACTTTATGAAGAAATATTTAAATAATGTAT	2924
Qy	2890	ACATCTCTAACGAGACTTCACTACACTTGTAATCTTAAGCCGATGATTCAACTCT	2949
Db	2925	ACATTTTATTAATGAATCTTTTACAAATTTTGTAAAACTTAAGCTGATGATATTTATTTCA	2964
Qy	2950	CTTAACGATGATCTTAACGATAGAAAGCTGGAAGAGACATCAATAGCTGAAGACACA	3009
Db	2985	TTGAATGATGATCAAAAAAGAAATTTAGAAAGATATTTAATTAATTAATAAAAACT	3044
Qy	3010	CTGCACTGAGCTTCGACCTGTACACACAGTACAAACTGAACCTGAGAGACTCTTGAC	3069
Db	3045	TTACGATTAATCTTTGATTTATATATTAATTAATTAATTAATTAATTAATTAATTAAT	3104
Qy	3070	AAGAAGACAGCTCGGACAGTATTAAGATGACATCAACAGCTGACCTGCTGCTCAAGAG	3129
Db	3105	AAAAAGAAAAAGTTGGTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3164
Qy	3130	CAGCTTGAAGACAACTCACTCACTGTAACATATCCGAAACAGCTACTGAGAACTTCA	3189
Db	3165	CAATTAGATCAAAATTTGAATTCACCTTAATTAACCCAAAGCATGTATTAACAAACTTTCT	3224
Qy	3190	GTTGTTCTTCAACAAAGAAAGAACCCGAGATGCGCGACAGACACACTGTGAGAAC	3249
Db	3225	GTTTTCTTTAACAIAAAAAAGAGCTGAATAGCGAAACGTAAGAACATTAATTAACAAAC	3284
Qy	3250	ACCAAGATTCTTCAACCACTCAACAAAGGCTGCTGACATTAATTAATGAGCAGCTTCT	3309
Db	3285	ACAAAATATTTATTGAAACATTTAAGAGCACTTGTAAATATTTATATATGATGATATCATCT	3344
Qy	3310	CCCTGGAAGACTCTCTCCGAGAGAGATCCAGCCGAGATTAAGTACGCCAGCCTCGAG	3369
Db	3345	CCATTAATAAACTTTAAGTGAAGATCAATTAACAAACAGAAATATTAATTAATTAATTAAG	3404
Qy	3370	AACCTCAAGCTCTGCTTAAGCTCGAAGCAGCTTAAGACAAACCTGAACTGGAGAG	3429
Db	3405	AACTTTAAGTATTAAGTAAATTAAGAGGAAATTAAGAGATTAATTAATTAATTAAGAAAG	3464
Qy	3430	AAGAAGCTCACTCTCTCTGACGGAGCTGATCACTGATGCGGAGCTGAAGAACTC	3489
Db	3465	AAAAATATATCATCTTATCAAGTGGATTACATCATTTAATTTGCTGAATTAAGAACATTA	3524
Qy	3490	ATTAGAACAAGAACTACACCGCCAAATAGCCCAAGCGAGAAATATATACAGCTGATTAAC	3549
Db	3525	ATTAATAATAATAATTTATACAGTAATTTCTCCAAGTGAAGAAATTAATACGATTTAACAT	3584
Qy	3550	GCACTGGAATCTTACAGAAGTTCTGCTGGAAGAACAGATGTGCCACTGTGTGCTCT	3609
Db	3585	GCATTAGCAATCTTCAAAAAATTTCTCCGAGAAAGCAAGATGTTGCAACAGTTGTAAGT	3644
Qy	3610	GAACTGCTCCGACACAGCAGAGCTGTCAACCTTAAGAACCTGCAATTAATCATATCTC	3669
Db	3645	GAAATGTGATCCGACACTTAAGAACAAAGTCAACCAAAAGAAACGAGCATCAACTATATGA	3704
Qy	3670	GGAGCCGAGTCAATTAATTAACACATCTCGAAGAGCTGAGAGATGAGCTGATGATGCTC	3729
Db	3705	GGAGCAGAGTCTTAACACATTAACACATATCAAAATATGCTGATGATGAAGTATGATGACGTA	3764
Qy	3730	ATCATTTGCTATCTTCCGCGAGAGCGAGAGAGACTACGATGACCTCGGCCAGGTGCTC	3789

Db 4197 TCTGATTTAATGCAATTTAAACATATATCTCAATGAAATGATTTGGAAGATTCATTT 4256
Oy 4060 AAGTTCTCAATTAAGACAGAGGATTAATTTCTGCTACTTACACTATATCAAGAC 4119
Db 4257 AAAATATGATTTGAGAACAAAAAACACACTTTTAAAAATCTTCAAAATATATAAAAGAA 4316
Oy 4120 TCCATGCGACCGGATTCATTAATTTGGCTTAAGTGTGGGTATTTACAAAGTCTGAGC 4179
Db 4317 TCAGTAGAAATGATATTAATTAATTTGACAGAGGATTAAGTATTAAGAAAAGGTTTAA 4376
Oy 4180 GAAAAATCAAGTCTGACCTTACTCTATTAATAAGTATATCAAGATAG----- 4230
Db 4377 GCGAAATTAAGATGATTTGAAATCAATTAATAAGTATCAAGAAGAAAAGAGGTTTC 4436
Oy 4231 -----CAAGCGAGAT 4242
Db 4437 CCATCATCACCAACAAACACCTCCGTCACAGCAAAAAGACGAACAAAAGAGAA 4496
Oy 4243 GAAAAATCTGCCCCCTTCTGTAATACATCGAAACCCCTGACAGACAGTCAACGACAAA 4302
Db 4497 AGTAAGTCTCTCAATTTTAAACAAACATGAGACCTTATCAATTAATGTTATATAA 4556
Oy 4303 ATGACCTCTGTAATTCACCTGAGGCGCAAGGCTCACTATACCTTACGAGAGAGC 4362
Db 4557 ATTGACGATTTACTTATTAATTAAGGCAAAAGATTAACGATTTGATGTTGAAAAAGAT 4616
Oy 4363 AATGGAAGTAAATCAAGAGCTGAACTCAACAAACATCAAGACAGCTGGCA 4422
Db 4617 GAAGCAGACTTTAAATTAACCTTAAGTATTAAGCAATTAAGTATTAAGCAAAATAGAT 4676
Oy 4423 GATTTCAGAAAAATTAACAATTTCTGCGAATTTGAGACCTGCTACCGATTTATTAACAC 4482
Db 4677 CTTTTTAAAAACCTGACGCTTGAACCAATTAATAATGATTAATGATGATGAGAA 4736
Oy 4483 AACATCTCTGACCAATTTCTGTCACCTGAGCTGTTGCGAAGAACCTGCCAAAAA 4542
Db 4737 AAAGATATGCTGGCAATTTACTTGTACAGAGATAG---TTCAAAATTTCTTAATACA 4793
Oy 4543 GTGCTGAGCAATTCCTGAGGCGCAACCTGCAAGGCTGCTGACATCTCCAGCACCA 4602
Db 4794 ATAAATCAAAATTAATTAAGGAAATTTCCAAAGATTTAAACATTTCCACACACCA 4853
Oy 4603 TGGCGAAGAAACAGTGGCCCCAGAAATAGCGCTGTTTCAGGACATCTGAGACGAGCGAA 4662
Db 4854 TGGCAAAAAACCAATGTCGCAAAATTTCTGATTTTTCAGACATTTTATGATGAAGAGAA 4913
Oy 4663 GAGTGCAGTGTCTCTGCACTCAACAAAGAGAGATTAAGTGGTGGAGAACCAAC 4722
Db 4914 GAATGTAATGTTTATTAATTAACAAAGAGAGTGAATTAATGTTGAAAATCCAAT 4973
Oy 4723 CCTACCTCAATGAACAAATGCGGGTGTGACGCCGATGCTTAATGACACGAGAGAAC 4782
Db 4974 CCTACTTCAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5033
Oy 4783 AGCGGCTTAAGCGAAATTAATCAATGCGAGTGTACTAAGCCGCGATCTTATCCATC 4842
Db 5034 TCAGGTACCAAGGAAATTAATCAATGATGATGATGATGATGATGATGATGATGATGAT 5093
Oy 4843 TTGACGAGGATTTTGTCTCCAGCTCTAATTTCTGGGACATCTCTCTGCTGATCTCT 4902
Db 5094 TTGATGATATTTTGTGAGATTTCTCTCAACTCTTAAGAAATATCAATCTTATTAATAC 5153
Oy 4903 ATGCTGATCTGTACAGCTTCAATCAATAGATGATGAT 4939
Db 5154 ATGTTAATATTAACAGTTTCAATTAATAAATGATAGG 5190

RESULT 10
AF062348 5243 bp DNA linear INV 09-FEB-2001
LOCUS AF062348
DEFINITION Plasmodium falciparum strain HNI merozoite surface protein 1 precursor (msp1) gene, complete cds.

ACCESSION AF062348
VERSION 1
KEYWORDS GI:3859842
SOURCE
ORGANISM Plasmodium falciparum.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
AUTHORS Jiang, G., Liu, R. Z., Daubenberger, C. A. and Pluschke, G.
TITLE Sequence analysis of the MSP 1 gene of Plasmodium falciparum from Haitian, China
JOURNAL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17
REFERENCE (5), 294-297 (1999)
AUTHORS Jiang, G., Liu, R. Z., Daubenberger, C. A. and Pluschke, G.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel CH-4002, Switzerland
FEATURES
Source
1..5243
/organism="Plasmodium falciparum"
/strain="HNI"
/db_xref="taxon:5833"
/chromosome="9"
/country="China"
/gene="msp1"
/gene="msp1"
/product="merozoite surface protein 1 precursor"
/note="binds spectrin"
/codon_start=1
/product="merozoite surface protein 1 precursor"
/protein_id="AAC7284.1"
/db_xref="GI:3859843"
/translation="MKIIIFLCSEFLPIINTOCYTHSYOELVYKLBALDEAVLTGYS
LFQEKVWLNSTGATVATTTSGSGSVTSGGGSAVASGSGSGSVAGSGS
RRTSPDSSSDSAKSYADLKHRYONVLTFTKEIKYELPELTHMLTLDNHGIFY
LIDYEELNELLYKNTFYDILRAKLNDACANSYCPFNLIKIRANLDELVYKIFYCY
RPLDNTIKDNGKMDYTKKRTITANIINLBSKRTIDONKANDEBKKIKYQAO
YNLPIYKQLOEAMNLISYLEKRLDTLAKNENIKLELIDIKITDKEKLTGSKPNP
LPENKKKEVEGHEKIKELIAKTIFENIDSLPTDLELEYLRKNKRVDPKSDPT
KSVQIDPKVYNGVIVPLPTDINSLAADNDKSYGGLAMPDKEKINKEITDNE
RKIPINIKKOIDLEKKNINTEKONKLEKDYKSKDYELLEKRYEMKFNPNPK
DYVDKISASRYTVVEKORYNNKSSSNYSYNNOKIKKALSTLYEDYSIKRGISEKDP
NHYTLTGLDEADIKKLTETEKSSEKNTLEKRFGLTHSANALESVDIYLYQYKYL
LIKRTIEDRLTEFLKNAQKDSIHVPITKPPKPEYIYLAKKEDEKIEIPY
KDLKKEQAVLSITOPLYAASSETTEGSGHTVYLSQGETELETVEGHTVITIL
PTPOSPPEKVEVYENSIEHKSNDNSQALFTVYLKRLDEFTKSYICHKILVNS
MDOKLLEYNTLPREKNKIKSCDPLDLIFNTONNIPAMYSLYDSMNDLOHLPFLYO
KEMTYLHKLEENHKKLEBQKQITGTSSTSPGATVTVTAOSATHSNOOSNA
SSTNPONVAAVSSGPAAVESHDPLVLSISNDLKGIVSLNLDNKRKVPPLITST
NTEYENKILANDTYFNDDIKQFVKSNSKVIITGLTEQKALNDEIKKLDITQLSF
DLNRYKLIKDLRLFNKKKELODKMOIKLLEKLEQLESKLSINSINPNHVNLFNSVE
KKKEAEIAETENTLENTKILKHYKGLVKNSESSPLTSESIOTEDNYANLEK
NRVSKIDGKIDNHLHGGKRLSFLSGISGLHITELKEVIRKRYTSPENNNKYN
EALSKYENFDEPAKVTVYTPPODVPYPSPLSVYSSSGSTRETOPTPSGSLTEL
QOVVOSONYDEEDSLVPLTEGSEDNDELDOVYGEALSTMDNLTLSGFENEYD
IYLPPLAGVNSLKQKLEKNTIFPNLNLNDLNSRKRKFLDVLBSDLMOQFHIS
NEYITIEDSKLINSKQNTLKSRYKESVENDIKRAQEBISYEVKLVAKYDLS
IKVYIKKEKEPSSPTPTPAPATDOKKEDYKPLFTNETLETLYNNLVNKIDYLMG
LKAKINDCNPVKDEAHYKTKLSDLKADIXIDLPKMTNDEATIKKILINDTKRDMG
KLUSTGLVONFPNTITISKLEGRQDMNLNISOHQCYKCKCENGSCRRHIDREBCK
LILNRQEDKCVENPNPCNNNGCGADACCTEDSDSGSKKITTCTKPTDSTPLFD
GIFCSSSFTLISPLILMLLYSPFI"

BASE COUNT 2303 a 677 c 686 g 1577 t
ORIGIN

Query Match 26.9%; Score 1331.2; DB 3; Length 5243;
Best Local Similarity 56.9%; Pred. No. 1.2e-281;
Matches 2919; Conservative 0; Mismatches 1978; Indels 231; Gaps 17;

[illegible]

Db	1128	AAACTATTAAATTAACTGATGATGTTTATTCTGATGATTCACCTGGAATTTGAATATTAT	1187
Qy	1144	TTGAGAGAGAGATAGAGATATAGACATCTCCGCAAAAGTCG-----GACAAAG	1194
Db	1188	TTTAAAGAGAAAAATAAAAAGTTGATGCTAACCTTAATACAGATCTAGCAAACT	1247
Qy	1195	GAAATACCCGAACCTAAATGAATATCCAGTGGTGAGGTACCCCTGCTTATTAACAT	1254
Db	1248	GTTCCAAATACCAAAGTCTCTTATTCCAAATGATATTGATATCTTTACCACTACTAT	1307
Qy	1255	ATCAACAA-----CGCTCTCAAGAGCTCAATAGCTTCGTGACTGATTAAACCC	1305
Db	1308	ATTCATTAATTCATTAGCTGCAGATATGATATAAATTCATATGTTAATGAATCCCT	1367
Qy	1306	TTCCGATTACGAAGAAACCCCTTAAGATATCTACACAGACATGAGAGAAAGATT	1365
Db	1368	GATACTAAAGAAAAATTAATGAAAAATTTATTCAGATTAATTAAGAAAAATATTC	1427
Qy	1366	ATCAACGAATTCAGAGAGATAGATCAAAATTTGAGAGAAATTAAGAGAGACAGAA	1425
Db	1428	ATTATTAACATTAAAAACCAATTGATTTAGAGAGAAAAACATTAAACACAAAGAA	1487
Qy	1426	AGTTACGAGAGACCGAGCAAAAAGTCTAAACGATATCTCAAGAGTATGAAAAGCTGTG	1485
Db	1488	CAAAATATAAAATTTACTGGACATTTATGAAAAAGTCAAAAAAGATTATGAAGATTA	1547
Qy	1486	AACGATCTATGATTCGCAAAATTCACATTAATCATGACCTGACCACTCGAGAAAA	1545
Db	1548	GAAAAATTTTATGAATGAAATTTAATATTAATTTTGACAAAGATGCTGATATAAATA	1607
Qy	1546	ATGGGAAAAAGCTACTCTTACAAAGTGGAGAAAGTACACACCATTAATCCTTTGCATCC	1605
Db	1608	TTCACTGCAAGATATACATATATATGTTGAAAAACAAGATATAATTAATTTTCATCC	1667
Qy	1606	TATGGAATTTCTAACATATATCTGAGAAAGCTACCAAAAGCTTTAAGTATATGAGAGAC	1665
Db	1668	TCTAATTAATTTGTTATATATATGTTCAAAATTTAAAAAGGCTTTTCATATCTTGAAAT	1727
Qy	1666	TATTCCTCAGCGACATTTGTTGGAGAAAGAACTAAAGTATTTACAAATCTCTAAGT	1725
Db	1728	TATTTCTTACGAAAAGATTTCTGAAAAAGATTTTAACTATTTATATCTTTGAAAACT	1787
Qy	1726	AAGATCGAAAAAGATCGAGAGCGCTTGTGGAACTTTAAGAAAGATGAAGACAGTTG	1785
Db	1788	GGCCTCGAAAGCTGATATATAAAAAATTAACAGAAAGATTAAGATGAGTGAACAAATTT	1847
Qy	1786	TTTGGAGAGAGATTTACAAAAGACAAAATTAACAGATGAGAGATCTCGAGGTCTCC	1845
Db	1848	CTAGAAAAAAATTTTAAAGACTTAACATCTAGGAAATG---CTTCTTGAAATATCT	1904
Qy	1846	GATATTTGTTAAAGTCCAGTGCAGAGGTCCTCATGACACAGATTGATGAATCAAG	1905
Db	1905	GATATTTGTTAAATTTACAGTACAAAAAGTTTATTATTTAAAAAATATGAAGACTTAAGA	1964
Qy	1906	AAGACTCAACTCTGAGAGAGCTGGAATTAACAATATATACATGTGCGAATAGT	1965
Db	1965	AAGATAGAAATTTATTTAAAAATTCACAACTAAAAAGATAGATTTCAATGCTCAAAATATTT	2024
Qy	1966	TATTAAGCAGAGATATAGAGAGAACCATACTCCATCTGTCCTCAAGAAAGATATAGC	2025
Db	2025	TATTAATACCAAAATTAACAGAACCATTTATTTAATTTGTTTAAAAAAGAGCTAGAT	2084
Qy	2026	AAACGTAAAGTGTGATGCGCCAAAGTTCAGAGCCTGATCAACGAGAGAGAAATTAAT	2085
Db	2085	AAATTTAAAAAGATTTATACCAAAATTAAGACATGTTAAA--GAAAGACAGCTGTCT	2142
Qy	2086	AAACTGAGAGAGCTCAGATTAATCTCGAGCCCTTCACAGAGAGAGATTAACGCGACG	2145
Db	2143	TATCAAGTATATCACACACTTTAGTTGACGACGAAGGAAACCACTGAAGAT--GGGGTGCAC	2201
Qy	2146	GCTACCAACAGCCGGAGACAGAGCCGGTTAGCTCTCGAAGGCGATATGCTGCAACT	2205

Db 2202 TCACACACATATATCCCAATCAGAGAAACAGAAAGAAAGAAACAGTAGACAC 2261
QY 2206 CAGGACACAGACAGACAGACAGCCTCGATGCGCAGTGCCCGCTTCAGAGCTAAA 2265
Db 2262 ACAACAAAGGTAACAAATACATTTACCACCAACACCAACCATCAACCAACAAAGTAATAA 2321
QY 2266 GCTCAAGTGCCTACACCAACAGCCTCTGTGAATTAACAAGACGAGAAATGTGCAACCTG 2325
Db 2322 GTGTGTAATAATTCATATGAAACATTAAGAGTAATGACAAATTCACAGCCTTGACAAAACA 2381
QY 2326 GACTACCTTGAGAACCTCTATGAGTTCTGTAATACATCTACATCTGCACAAATATATC 2385
Db 2382 GTTATATCAAGAAATTTAGATGAATTTTAACTAAATCATATATATGTCATTAATATATAT 2441
QY 2386 CTCGCTCTCAGACGACATATGACAGAGAAGATTTCTTAAACAGTACAAAGTAAACAGAA 2445
Db 2442 TTAGATTAACAACTAGATATGACCAAAATTTATAGAGTAATTAACCTTACTCCAGAA 2501
QY 2446 GAGGAGATAAAGTCTCTCTGTGATGACGACCTGCTGTCATATCCAGAACAA 2505
Db 2502 GAAAAAATGAAATTAATTCATGATGATCCATTTAGATTTATTAATTAATTCAAAATTAAC 2561
QY 2506 ATTCCTGTTATGATATTTATGTTGATAGCTCAACAAATCTCTCTCAACTGTTTCAATG 2565
Db 2562 ATACCTGCTATGATATTTATATATATGATAGTATGACAAATGATTTTACAACATCTTTT 2621
QY 2566 GAGATATATGAGAAAGATATGCTGCAACCTGTAATTAACCTCAAGACAGCAAGAT 2625
Db 2622 GAATATATCAAAAGAAATGATTTATTTATTTACATTAACCTAAAGAGAAATATCAATC 2681
QY 2626 AAGAACTTCTGAGG----- 2641
Db 2682 AAAAAATTTATAGAGACCAAAACAAATACCTGAGACATCATACATCCAGTCTGGA 2741
QY 2682 -----AACCTAAGAGGCTCTCAGCTCTGTTTAAACCTCTCTTCAGCTCCAGCAAC 2696
Db 2742 AATACAAACCGTAATATACGCTCAATCCCAACTCAGACTAATTTCCCAAAACCAACAACTCA 2801
QY 2697 ACTGCTCTCAGACCTCAAGCA----- 2730
Db 2802 AATGATCTCTACCAATACCCAAATGAGTATGCTATCATCTGCTCTGCTATGTT 2861
QY 2721 AGCGTAAAGACGACACTCTCATCGACCAACCTTAATACCTCACTGAAACCTGTTAG 2790
Db 2862 GAAGAAATCATGATCCCTTAACAGTATGCTATTTAGTAAGATTTGAAAGTATGTT 2921
QY 2791 AACATCCCTGCTCTCGGCAAGATTAAGAACT---CTACCAAGAACTTATGACAGAA 2847
Db 2922 AGTCTCTTAATCTTGAAATTAACCTAATCCATTAACCAATTTCTTACAACA 2981
QY 2848 TCGTCCGAACTCTTACGAGAGATGAGTAAGAGACAGACACATCTATTAAGAGAGC 2907
Db 2982 GAGATGAAAAATTTTATGAGATATTTTAAAAATTAATGATACCTATTTTATATGAT 3041
QY 2908 TTCCTAACTTCTGTAATCTAAAGCCGATGATATCACTCTTACGATGATCTTAA 2967
Db 3042 ATCAACAATTCGTAATCTAATCAAAAATTAATTAAGATTTACAGGTTGACC---GAACACAA 3098
QY 2968 CGTAAGAGCTGGAAGAGAGATCAATTAAGCTGAAGAGACACTGCAACTGAGCTTCGAC 3027
Db 3099 AAAAAATGATTAATGATTAATTAATAAATTAAGATTAATTAATTAATTAATTAATTAAT 3158
QY 3028 CTGTACACAAAGTACAACTGAACTGAGAGACTCTTCGACAAAGAAAGACACTGCGC 3087
Db 3159 TTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3218
QY 3088 AAGTATTAAGATGACAGATCAAGAAATGATCTGCTCAAGGAGACCTTGAAAGCAAACTC 3147
Db 3219 CAAGCAAAATGCAATTAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3278
QY 3148 AACCTACGAAACATCCGAAACACTACTGCAAGAACTTCTGCTTCTTCAACAAGAAAG 3207
Db 3279 AATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3338

QY 3208 AAGAAAGCGAGATGCGGAGACAGAAACACTCTGGAAGAACCAACCAAGATTTCTTCAAA 3267
Db 3339 AAAGATGTAAGAAATGCGAAACATTAAGAAACACATTAAGAAACCAAAATTAATTAATTA 3398
QY 3268 CACTCAAGAGGCTGCTGAAGTATTAATGAGCGAGTCTTCTCTGAGACCTCTGCC 3327
Db 3339 CATTTAAAGGACTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3458
QY 3328 GAGGAGACATTCAGACCGAGAGATTAATGACGACCTCTGAGAACTTCAAGCTCTGCT 3387
Db 3459 GAAGTATCAATTAACAAAGAGATTAATGATCCATTTAGAAAAATTTAGAGTATTAAGT 3518
QY 3388 AAGCTCGAAGGCAAGCTGAAGACACTGAACCTGGAAGAAAGAAAGAACTGAGCTACCTC 3447
Db 3519 AAAATGATGGAAGAACTAATGATTAATTTATTAAGAAAGAAAAATTTATCTTTT 3578
QY 3448 TCTAGGAGACTGACATCACCTGATCGCGAGCTCAAGAGATCATTAAGAACAGAACTAC 3507
Db 3579 TCAATGATTAATCATCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3638
QY 3508 ACCGCAATAGCCAGCGAGATTAATACAGAGTGAATTAACGACCTGGAATCTTACAG 3567
Db 3639 ACAGTAAATTCCTCAAGTAAATTAATTAAGAAAGTTAACGAAGCTTTAAATCTTACGAA 3698
QY 3568 AAGTCTGCGCTGAGGAAGAGATGCG----- 3596
Db 3699 AATTTTCTCCAGAAAGAAAGTTTCAACAGTTGTAACCTCAACCTCAACAGATTAATCT 3758
QY 3597 CACTGTGCTGCTGAATCTGCTCCGACACACTGAGGAGCTCAACCTAAGAA----- 3650
Db 3759 CCATCTCATTAATCTGTAAGGTAAGTGTAGTTCAGATTCACAAAGAAACACAA 3818
QY 3651 --GCTGATCTACTCATGCTGGAGCCGAGTCCAAATTAATTAACAACTTCCAGAACTG 3708
Db 3819 ATACCACTTCAGGCTCTTATTAACAGAAATTAACAAAGTAAGTAACAAATCAAAATTAAT 3878
QY 3709 GAGGATGAGTGCATGAGCTCATCTGCTATCTTCCGCGAGAGGAGGAGACTAC 3768
Db 3879 GACGAAAGAGATGATCTCTGATGTTATTAATCCATTTTGGAGAAATCCGAAGATTAATGAC 3938
QY 3769 GATGACCTGCGGAGGCTGACCGGTGAGGCTGCTCTCTGCGATTTGATTAAT 3828
Db 3939 GAATATTTAGATCAAGTATGTAATCTGGAAGC---AATATGCTGACATGATTAATATC 3995
QY 3829 CTGTCCAAATTCGAAAGATTAAGAGTGTCTATGGAACCTCTGCGAGGCGCTAT 3888
Db 3996 CTCTGAGATTTGAATGAATGATGTTATTAATTAATTAATTAATTAATTAATTAATTAAT 4055
QY 3889 AGGCTCTCAAGAAACAGCTGAGAAATTAAGGATGACCTTCAATGTCACAGTGAAGAC 3948
Db 4036 AGAAGCTTAATAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4115
QY 3949 ATTTGAACAGCGCTTAAATTAAGAGAAATTTCAAGACGCTGAGAGCGACTTG 4008
Db 4116 ATCTTAATTAATCAAGCTTAAGAAAGAAATTAATTTCTAGATGATTAATTAATTAATTAAT 4175
QY 4009 ATTCCTTAATTAAGACCTGACCTCTTAAGTATGCTGCAAGGAGGAGGAGGAGGAGGAGG 4068
Db 4176 ATGCAATTTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4235
QY 4069 AATAAGAGAGAGGAGTAATTTCTGCTAGTTTCACTATATCAAGAGCTCACTGAC 4128
Db 4236 AATTCAGAAACAAACAAACACTTTTAAAGATTAATTAATTAATTAATTAATTAATTAATTAAT 4295
QY 4129 ACCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4188
Db 4296 AATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4355
QY 4189 AAGTCTGACCTGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4230
Db 4356 AAGGATGATTTAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4415


```

Qy 4231 -----CAGGCGAGAAATGAAATAT 4251
Db 4416 CCACCAACAACCTCCGTCACCAGCAAAAACAGACAGCAAAAGAAAGAAAGTTC 4475
Qy 4252 CTGGCCCTCTGATTAACATCGAAACCTGTACAGACAGCAAAATGACCTC 4311
Db 4476 CTTCACATTTTAAACAACATGAGACCTTATACAAATTAAGTAAATTAATGACAT 4535
Qy 4312 TTGCTAATTCACCTGAGGCGCAAGTCCTCACTATCTTACGAGAAAGCAATGTGAA 4371
Db 4536 TACTTAATTAAGTAAAGCAAGATTAAGCATTTGAATGTTAAAGATGAGACAT 4595
Qy 4372 GTTAAATCAAGAGCTGAATACCTCAAAACAATCCAGACAGCTGCGCATTTCAAG 4431
Db 4596 GTTAAATTAAGTAAATTAAGTAAAGCAATTAAGTAAAGTAAAGTAAAGTAAAG 4655
Qy 4432 AAAAATACAAATTTGCTGGAATTCGAGACCTCTCTACGATATATACCAACATCTC 4491
Db 4656 AACACTAAGCACTTGAAGCAATTAATAAATTAATTAATTAATTAATTAATTAATTA 4715
Qy 4492 CTGACCAAGTTTCTGTCACCTGTCATGTCGTAACCAACCTGCGCAACAGTGTAGC 4551
Db 4716 CTGGCAAAATTAAGTAAAGCAAGATTAAGTAAATTAAGTAAATTAAGTAAATTA 4772
Qy 4552 AATCTGCTGAGCGCAACCTGACGAGCTGACATCTCCAGACCAATGCTGAAG 4611
Db 4773 AATTAATTAAGCAAGAAATTCAGATATGTTAAACATTTTCAACACCAATGCTGA 4832
Qy 4612 AAAACATGCCCCAGAAATAGCGGCTTTTCAGGACATCTGAGACGCGCAAGTGCAG 4671
Db 4833 AAAACATGCTCAGAAATTCGATGTTTCAGACATTTTACATTAAGAAAGAAATGAA 4892
Qy 4672 TGTCTCTCACTCACTCAACAGAGAGATTAAGTGGTGGAGAAACCAACCTTACTGC 4731
Db 4893 TGTCTCTCACTCACTCAACAGAGAGATTAAGTGGTGGAGAAACCAACCTTACTGC 4731
Qy 4732 AATGAACAAGTGGGCGGTGTGACGCGCATGTAATTCACGAGAGAGAGAGAGAG 4791
Db 4953 AATGAACAAGTGGGCGGTGTGACGCGCATGTAATTCACGAGAGAGAGAGAGAG 4812
Qy 4792 AACGGAAGAAATTCAGATGCGATGACGAGCGGACCTGATCAGCTTTCGAGAG 4851
Db 5013 AACGGAAGAAATTCAGATGCGATGACGAGCGGACCTGATCAGCTTTCGAGAG 4872
Qy 4852 AATTTTCTGCACTCACTCACTCACTGCGATCTCTCTCTGATTCCTCATGCTGAT 4911
Db 5073 AATTTTCTGCACTCACTCACTCACTGCGATCTCTCTCTGATTCCTCATGCTGAT 4932
Qy 4912 CTGTACAGCTTCACTCACTCACTGATG 4939
Db 5133 TTATACAGTTTCACTCACTCACTGATG 5160

```

RESULT 11
AF218248 5085 bp DNA linear INV 24-JAN-2000
LOCUS AF218248
DEFINITION Plasmodium falciparum major merozoite surface antigen (gpi195) gene,
complete cds.
ACCESSION AF218248
VERSION AF218248.1 GI:6740099
KEYWORDS
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE 1 (bases 1 to 5085)
AUTHORS Shan, Z. X., Yu, X. B., Li, X. R., Ma, C. L. and Fang, J. M.
TITLE Molecular cloning and sequence analysis of major merozoite surface
antigen (gpi195) gene of Plasmodium falciparum isolate FC01/HN
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5085)
AUTHORS Shan, Z. X.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Parasitology, Sun Yat-sen

```

FEATURES
source University of Medical Sciences, 74 Zhongshan Road II, Guangzhou,
Guangdong 510089, P.R.China
Location/Qualifiers
1..5085
/organism="Plasmodium falciparum"
/isolate="FC01/HN"
/db_xref="taxon:3833"
/country="China"
1..5085
/gene="gpi195"
1..5085
/codon_start=1
/product="major merozoite surface antigen"
/protein_id="AAF27526.1"
/db_xref="GI:6740100"
/translation="MKIIIFLCSPLEITINOCYTHESYOELVRLKLEALEDAVLGYS
LFQEKMYLNEGTSGTAVTTSTPGSGSVTSFGSGSVASVAGSGSGSVAGSGS
RTNPSDSNSDASVADLKRVONYLTIELYPELEDLTNHMLTLCNIDHGRKY
LIDYEELINELLYDLRAKLNACANSYQIPENLIRANEDLVKRLIVGY
RPLDNKDNVGMEDYIKKNTTIANINELIEGSKTIDOKNADNEGKRLVQAO
YLFYINRLOEAHNLISYLERIDPLKKNENIKLEIDIKTDAEKLTSKPNP
LEPNKKVEGHEEKIKETAKIKENIOSLFPDPLELEYLREKKKVVVPEKSDPT
KSVQIPKPYPNGLIYPLPUDIHNSLADNKNYSGLMDNPELEKINEKITDNKE
KRIFLNINIKQIDLEKNTNHRKQKLEBDEYSKDYELERFEMKNNFNDK
DVYDKIFESARYTYNVEKOKYNNKFSNNNSVYVOKLKALSYLEDYSLRGIISKDF
NHVYLTGTELEADIKLEIEISENKKILEKNFKGLTHANSLSVLDVKIJOVKV
LKIKLEIDELFLPNAOKDSIHPNLYRKPNRPEPYLLYLKREVDKLEIPKY
KMLKKEQAVLSITOPLYAASETEDEGSHRNTLSQSETEETEYEGHTTAVITL
PTQPSPREKVVVENSIEHKNDSQALTKVYLAKIDELFTKSYICHKYLIVNS
MOOKLEVINLPEERNEKLESCDPLDLNITONNIPAMSLDSMNDLQHLFPELQ
KMIYLLHLEKNEHKLLEBOKOITGSSSPGNTVNTAOSATNSNOSQNA
SSTNONGVAASGPAVVEESHDPPLVLSISDLGIYSILMLGKTKVNPDLTST
EMEKRYENLNNDYFNDIKDOPKNSKRTLTGLEDKNLNINNPHVLDNFSVF
DIYNKYLKLDLFPKKRELGDOKMOKITLTKLEKLESLNINNPHVLDNFSVF
NKKRAEIAETENTLNTLILKHKGLKYVNGESSPKTISEVISOEDNVALEK
FVLSRDKLDNDNLHGRKLSFSSGLHLITELKEVITKNTTGNPSNNKVN
EALKSYENFLPAKVTYVTPQPVTPDSVRSVSGSGSTKEEFOPTSCSLTEL
QOVOSONDEDEDSLVLPPIGESDENDEYLDQVYTGAIISVTMDNISGENEYD
IYKPLAGYRSILKOKIENITTPNLNIDNSRLKRRYLDVLESDLMQFKHS
NHYIIEDSFKLNSBOKNTLKSYSIKRSVNDIKFOEGISYIEKVLAKYKDDLES
IKVYIEKEKEPSSPTTPSPARKDEOKKEKELPDLTINFTLNINNYKIDYLLIN
LKAINDCNEKDEKHAHVKTITLSDLAIDOKIDELKNTNDEFAIKLINDTKMGL
KLSTGLYONFPNTIISKLEIFQDMNLSOHQVKKOCPENSGCFRLDREBCK
LINVROEGDKVEENPNCENNGGCDADAKCTEBDSGNGKRTICECTKPSYPLFD
GIFCSSNGLISPLIIMLILYSFI"
BASE COUNT 2244 a 671 c 678 g 1492 t
ORIGIN
Query Match 26.9%; Score 1330.8; DB 3; Length 5085;
Best Local Similarity 57.0%; Pred. No. 1,5e-281;
Matches 2915; Conservative 0; Mismatches 1972; Indels 231; Gaps 17;

```

Qy 10 ATGAATAATCATTTTCCCTGCTCATTTCTTTTATCAATCACTCAGTGGT 69
Db 1 ATGAAGTCAATTTCTTTATGTTCATTTCTTTTATTAATAATACAAATGTGA 60
Qy 70 ACCCAGGAATCTATCAGAGGCTGTGAAGAACTTGAAGATGCCGCTCT 129
Db 61 ACACATGAAGATTAAGCAAGACTTGTCAAAAACCTAGAACCTTAGAAGTGCAGATTTG 120
Qy 130 ACCGATACAGCTGTTCAGAGAGAGAGATGCTGTGANTAGAGGACGAGTGCAG 189
Db 121 ACAGGTTATGTTATTTTCAAGAGAGAGAGATGTTAATAAGACACAACTGAGACA 180
Qy 190 GCCGTACACACGACACCCGCTTAAAGGCTGTGCTAGCGGTGCTCCGCTGG 249
Db 181 GCTGTTACAACTGATACCTGCTGAGTTACGTTACAGTGGTTAGGTTG 240
Qy 250 TCTGTGGCCTCGGGGGTTCGCTCGGCTCGGCGGACGCTGACATCAGTGCAGT 309
Db 241 TCAGTTGCTTCACTTCTCTC-----AGCTGTTACAGTGGCTCAGTT 282

Ddb	2437	GA	AAAAATGAAATTA	AAAAATCATGATGC	CAATTCGATTTTATTTAATTAATTC	AAATATAC	2438					
Oy	2506	AT	CCCTGATATG	ATCTGATGTC	GATAGCC	TACACAAATTCCTCTCA	ACTGTTCA	2505				
Ddb	2497	AT	ACCTGCTATG	ATATATATAT	ATATAGATG	ATACAAAGATTTTAC	CAATCTCTTTT	2556				
Oy	2566	GAG	ATATATGGA	GAGATGGCTG	CAACCTAT	TAACCTCA	AAAGACAGCAAGAT	2625				
Ddb	2557	GA	ATTAATTC	AAAGAAATG	ATTTATTA	TTTACATTAAC	CTAAAGAGAAATAT	CAATC	2616			
Oy	2626	AAGA	CTTCTG	GAGAAAG	CTAAGAAG	GTCTC-----	-----	2657				
Ddb	2617	AAAA	AAATTAAT	TAGAGAG	CAAAAAC	AAATTAAC	GGACAATCACTAC	ATCCAGTCTG	2676			
Oy	2658	-----	-----	CA	CTCTGTTAA	ACTCTCTCTT	CCAGCTCAT	CAAC	2686			
Ddb	2677	AA	TACAA	CCGTAAT	TACTGCTCA	TCCGCACT	CA	CGATTAATTC	CCAAAAC	CAATCA	2736	
Oy	2697	AC	GTCTCT	CACAC	CTC	TAAGACA	-----	-----	GGCC	CAAGT	2730	
Ddb	2737	AA	TGATCT	CTC	CAATTA	ACC	CAAAATGG	ATAGTGT	ATCAT	CTG	GTCTG	2766
Oy	2731	AG	CC	TAC	CGAC	CACTCT	CAC	TGACCA	CTTAAT	TA	CTCACT	2790
Ddb	2797	GA	AAAG	AGTC	ATG	ATCC	CTTAAC	AGATTTG	TAAGATTTG	TAAGATTTG	TAAGATTTG	2856
Oy	2791	AA	CACTCT	GTCTC	CGC	CAAGATTA	AGACAT---	CT	ACCA	AGAACTTA	TTGAC	2847
Ddb	2857	AG	CTCTTA	ATCTT	GGA	AAATTA	AC	TAAGATTTG	TAAGATTTG	TAAGATTTG	TAAGATTTG	2916
Oy	2848	TC	GTCC	GGA	ACTT	CT	CA	GGAAGAT	CTGA	AAAGAC	CACTCT	2907
Ddb	2917	GAG	ATG	GA	AAAAATTTT	TATG	GAATATTTT	AAAAAA	TAAGATTTT	TAAGATTTT	TAAGATTTT	2976
Oy	2908	TT	CACT	AACT	CTCG	GAATCTTA	AA	GGCGAT	GATATCA	CTCTTA	ACGATTA	2967
Ddb	2977	AT	CA	ACAT	TTCTG	TAATTT	CTAA	AGTTAC	AGTTTAC	AGTTTAC	AGTTTAC	3033
Oy	2968	CG	TAGA	AGCTG	GAG	AGACAT	CAATTA	AGCTG	A	GAGAC	CACTG	3027
Ddb	3034	AAAA	ATGATTA	TATGATTA	AAAAATTTT	AAAAAA	TAAGATTTT	TAAGATTTT	TAAGATTTT	TAAGATTTT	TAAGATTTT	3093
Oy	3028	CT	GA	CAAC	ATGCA	AACTGA	AAACG	AGAGCTCT	TG	ACAA	GAAGAC	3087
Ddb	3094	TT	AT	TAATTA	ATTAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	3153
Oy	3088	AA	GT	TAAGAT	GCAGAT	CAAGAT	GTGACTCT	GT	CA	AGAG	CACTG	3147
Ddb	3154	CA	MA	CAAAAT	TAATTA	AAAAAA	CTTACTTTT	TTT	AAAA	GAAC	CAATTA	3213
Oy	3148	AA	CT	CA	CAAT	TCGGA	AAACG	ATGAC	TA	CTGAT	CTGCA	3207
Ddb	3214	AA	TT	CTTA	TAATTA	CCCACTTA	ATTTG	TAATTT	TAATTT	TAATTT	TAATTT	3273
Oy	3208	AA	GA	CCG	AGATCG	CGC	GAGAC	AGAGAC	CTCTG	GA	ACCAAGATTT	3267
Ddb	3274	AA	AG	AG	AG	AG	AG	AG	AG	AG	AG	3333
Oy	3268	CA	CT	CA	AAAG	CCCT	CTG	CA	ATTT	AT	TATG	3327
Ddb	3334	CA	TT	AT	TAATTA	ATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	3393
Oy	3328	G	AG	CA	AT	TCG	CA	AG	CA	AT	TCG	3387
Ddb	3394	GA	AT	CA	ATTT	CA	AAAGATTA	TTATTT	TAATTT	TAATTT	TAATTT	3453
Oy	3388	AA	GT	CA	AG	CA	ATTT	CA	ATTT	CA	ATTT	3447
Ddb	3454	AA	AT	TA	AT	TA	ATTT	TA	ATTT	TA	ATTT	3513
Oy	3448	TC	TA	AG	CG	AT	GC	AT	GC	AT	GC	3507
Ddb	3514	TC	AA	GT	AT	TA	ATTT	TA	ATTT	TA	ATTT	3573

QY	3508	ACGGCACTACGCCAACCGGAGTATATACAGACGTCGAATTAAGCGCTGGATCTTACAG	3567
Db	3574	ACAGGATATCTCCAAAGTGAAATTAATAGAAAGTTTAACGAAAGCTTTAAATCTTAGCA	3633
QY	3568	AAGTTCCTGCCTGAGGAGACAGATGTCGC-----	3596
Db	3634	AATTTTCTCCGAGAACGAAAGTTTACACAGTTGTAACTCCACCTCAACGAGATTA	3693
QY	3597	CACGTGCTGTCTAATCTGCTCCGACACACTGGAGCGTCTCAACTAAGA-----	3650
Db	3694	CCATCTCCATTATCTGTAAGGGTAAGTGGTAGTTCAGGATCCACAAAAGAAACACAA	3753
QY	3651	--GCCTGATCTACTCTATGTCGGAGCGAGCTCCAAATACATTAACCAATCTCAGAAC	3708
Db	3754	ATACCACTTACGAGCTCTTTATTAACAGATTTACACAAGTAGTACATCCAAATATAT	3813
QY	3709	GACGATGAGTCGATGACGTCATCATTTGCTATCTTCGGCGAGAGCGAGAGCATAC	3768
Db	3814	GACGAGAGAGATGATCTTCCTAGTTGTAATACCATTTTGGAGAAATCCGAAGATTA	3873
QY	3769	GATGACCGCGCCGAGTGGTCACCGGAGGCGCTGACTCTCCGTGATGATTAAT	3828
Db	3874	GAATTTTAGTCAAGTAGTAACCTGGAGACATATCTGTACAAATG---GATATATC	3930
QY	3829	CTGTCAAAATCGAAGACGATACGAAGTGCCTATCTGAACCTCTGGACGCGCTAT	3888
Db	3931	CTCTGAGATTTGAAATGGAATGATGATGATTAATTTTAAACCTTTGCTGAGATATAT	3990
QY	3889	AGGTCTCGAAGAACGCTGGAGAAATTAACGTATGATCCTCAATGTCAACGTGAAGAC	3948
Db	3991	AGAGCTTAAAAAACAATTTGAAAAAACATTATTTACATTTAAATTTAAATTTGAACAT	4050
QY	3949	ATTCTGAACAGCCCTTTAATTAAGAGAAATTTCAACAGCTGTGGAGCGACTTG	4008
Db	4051	ATCTTAATTCAGCTCTTAAGAAACGAAATATTTCTTGATGATGATTAATCTGATTTA	4110
QY	4009	ATTCCCTATTAAGCTGACCTCTCTACTACGTTGTCAAGACCCATACAGTTCTC	4068
Db	4111	ATGCAATTTAAACATTAATCTCAATGATATCATTTATTTGAAGATTCATTTAAATATG	4170
QY	4069	AATAAGAGAGAGAGATTAATTTCTGTATACATATATTAACGACTCCATCGAC	4128
Db	4171	AATTCAGACAAAAAACAACACTTTTAAAAAGTTCAAATATATTAAGAAATCATGTA	4230
QY	4129	ACCGATATCAATTCGCTAATGATGTGTCGGGTATATACAGATCCTGAGGAAAAATAC	4188
Db	4231	AATGATATTAATTTGACAGGAAAGTATTAAGTTATTAAGAAAGTTTGAACGAAATAT	4290
QY	4189	AAGTGTGACCTGTGCTATTTAAAAAGTATATCAAGATAG-----	4230
Db	4291	AAGGATGATTTAGATCAATTTAAAAAGTTATCAAGAAAGAAAGAGTCCCATCATCA	4350
QY	4231	-----CAAGCGAGATGAAAAATAT	4251
Db	4351	CCACCAACAACCTCGCTACACGAAAAACAGAGAACAAAGAAAGAAAGTAAATGTC	4410
QY	4252	CTGCGCTTCGATATACATCAAGAACCCGTGTACACACAGTGAACGACAAATCGACCTC	4311
Db	4411	CTTCATTTTTAAACAAACATTTGAGACCTTATACAAATTAACCTTGATTAATTAATGACAT	4470
QY	4312	TTTCGTAATTCACCTGGAGGCGCAAGTGTCTCACTATTAACGAAAGACAAATGTGGA	4371
Db	4471	TACTTTAATTAACCTTAAGGCGAAAGATTAACGATTTGATGTGTTGAAAAAGATGAAGCAT	4530
QY	4372	GTTAAATCAAGAGCTGAACCTACCTCAAAACAATCAAGACAAAGCTGGCAGATTTCAAG	4431
Db	4531	GTTAAATTAACCTTAACCTTAGATTTTAAAGCAATGATGACAAATATGATCTTTTAA	4590
QY	4432	AAAAATTAACATTTTCCTCGGAATTTGACAGACTGTCTACGATTTATTAACACACAAATCTC	4491
Db	4591	AACACTTAACGCTTCGAAGAAATTTAAAAATTTGATTAATGAATGATGAAGAAAAAGATATG	4650

Oy	4492	CTGACCAAGTTTTCGTCCACTGGCATGGTGTTCCAAAACCTCGCCAAACAAGTGTGAGC	4551
Db	4651	CTTGGCAAATTAATTACTAGTAGAGATTAA--TTCAAAATTTTCTTAATAATATCA	4707
Oy	4552	AATCGCGTCGACGGCAACCTCGAGGGCTGCTGAACATCTCCAGCACCAATTCGTGAG	4611
Db	4708	AAATTAATTGAAGAAAAATTCAMAGATATGTTAAACATTTTACACACCATTGGCTAANA	4767
Oy	4612	AAACAGTCCCCCAGAAATAGCGGCTGTTTCAGGCATCTGACGAGAGCGGGAAGTGCAN	4671
Db	4768	AAACAATGTCAGAAAATTTCTGATGTTCAGACATTAGATGAAGAAGAAATGTMAA	4827
Oy	4672	TGTCCTCCGAACTACAAACAAGAGAGATTAAGTCGCTGGAGAACCCAAACCTTACCTGC	4731
Db	4828	TGTTTATTAAATTCMAAACAGAAAGGTGATTAATGTGTGAAAAATCCAAATCTTACTGT	4887
Oy	4732	AATGAACAATAGGGGGGTGAGCGCCGATCTTAATCACCAGGAGAGACAGCGGCTCT	4791
Db	4888	AACGAAATTAATGTGTGATGTGATGATGACATGCCAATTTACCGAAGAGATTCAGTATGC	4947
Oy	4792	AACGGAAGAAATTCATATGCGAGTGTACTAAGCCCGACTCTTATCCACTCTCGACGG	4851
Db	4948	AACGGAAGAAATTCATATGGAATGTACTAAGCTGATCTTATATCCACTTTTGATGTG	5007
Oy	4852	ATTTTTCTCGACGTCTAATTTCTGTGGCATCTCTTCGCTGATCTCATGTGATC	4911
Db	5008	ATTTTCTCGAGTCTCTCACTTCTTACGAATATCATTTATTAAATACATGTTATA	5067
Oy	4912	CTGTACAGCTTCATCTAA	4929
Db	5068	TTATACAGCTTCATCTTAA	5085

RESULT_12
PEP190G1

LOCUS	PEP190G1	5392 bp	DNA	linear	INV 03-JUN-2002
DEFINITION	Plasmodium falciparum p190 gene for surface antigen precursor.				
ACCESSION	X05624				
VERSION	X05624.1	GI:4186080			
KEYWORDS	antigen; glycoprotein; p190 gene; signal peptide; surface antigen.				
SOURCE	malaria parasite P. falciparum.				
ORGANISM	Plasmodium falciparum				
REFERENCE	1 Tanabe,K., Mackay,M., Goman,M. and Scaife,J.G. Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum J. Mol. Biol. 195 (2), 273-287 (1987)				
JOURNAL					
MEDLINE	88011243				
PUBMED	3079521				
REFERENCE	2 (bases 1 to 5392) Tanabe,K. Direct Submission Submitted (21-JAN-1999) K. Tanabe, Osaka Institute of Technology, Biology, 5-16-1 Ohmiya, Asahi-ku, Osaka, 535-8585, JAPAN Sequence revised by author On Jan 26, 1999 this sequence version replaced gi:9926.				
REMARK					
COMMENT	Location/Qualifiers				
FEATURES	1..5392				
Source	/organism="Plasmodium falciparum" /strain="MAD20" /db_xref="taxon:5833" /clone_lib="(lambda)NM149 genomic" 136..5241 /note="p190 precursor" /codon_start=1 /protein_id="CNA29112.1" /db_xref="GI:4186081" /db_xref="SWISS-PROT:P08569" /translation="MTKIFPLCSPIFLFICIMOCVTHESYOELVKILFALDAVLTYGS LFQEKWLNBEIGTSGLTAVYTTSTPGSSSVISGGVAVASVASGSGSVASGSGNS RIRNPSSSDSNRKTYADLKHRYQNLTFTLEIKLYEQLDNLNHLTLKNDGFY LIDYEELINELTKYNFYDDLPRKLNDACANSVCQIPFNFKLRANLIDLKLVFGY"				

Query Match	26.1%	Score 1290.4	DB 3	Length 5392
Best Local Similarity	56.4%	Pred. No. 1.1e-272		
Matches 2904	Conservative	0	Mismatches 1956	Indels 246
			Gaps	17
10	ATGAATCAATCTTCTCCTGCTTCAATTCGTTTATATCAATCAATCAATGAGCGCTG	69		
136	ATGAAGATCAATATCTTTATTAATGTCATTCCTTTTATTTATTAATAATACACATGTGTA	195		

Dh 2335 GAAACAGTAGACACACAACGCTAACATTAACATTACCAACCAAAAGAAATCAGCA 2394
Oy 2256 -----AGAGGCTAAAGCTCAGTGCCCTACACCACGCTCGTGATTAACAAGCCGAG 2310
Dh 2355 CCAAAAGAGTAAGATGTTGAAATTCATAGAACATTAAGATTAAGCAATTCACAA 2454
Oy 2311 AATGTCAGCAACCTGAGCTACCTTGAGAGGCTCTATGATTCCTGAAATACATCTACATC 2370
Dh 2455 GCGTTGACAAAACAGTTTATCTTAAGAATTTAGATGAAATTTTAACTAAATCATATATA 2514
Oy 2371 TGCCACAAATATATCTCGTCTCTCACAGCATATGAACGAGAGATTTCTTAACAGTAC 2430
Dh 2515 TGCTAAATATATTTAGTATCAAACTCTAGTATGAGCAAAAAATTTATTTAGAGSTATAT 2574
Oy 2431 AAGATAACCAAGAGAGAGATTAACGTCTGCTGTAATCCCTGAGACTGCTGTC 2490
Dh 2575 AATCTTCTCCAG 2634
Oy 2491 AATATCCAGAACACATTCCTGCTTATCTATCTATCTATCTATCTATCTATCTATCT 2550
Dh 2635 AATATCCAAATTAACATACCTGCTATCTATCTATCTATCTATCTATCTATCTATCT 2694
Oy 2551 TCTCACTGTTCTATGAGATATATGAGAAGAGATGCTGCAACCTGTATTAACCTCAA 2610
Dh 2695 CAACATCTCTTTTGAATTTATATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2754
Oy 2611 GACACGACAGATTAACACCTCTGAGG----- 2641
Dh 2755 GAGGAGAAATCAATCAAAAAATTTATAGAGAGCAAAAAAATTAACGTGAACATCATCT 2814
Oy 2642 -----AAGCTAAGAGGTCCTCAGCTCTGTTAAACCTCTCTCTTC 2681
Dh 2815 ACATCCAGTCTCGGAATACCAACCTTAATCTGCTCAATCCGCACTCACAGTATATCC 2874
Oy 2682 CAGCTCCATGCAACCATCTGCTCTCACACCTCAAGACAA----- 2720
Dh 2875 CAAAACCAACATCAAAATGATCCCTCTACCAATCCCAAAATGGTGTAGCTATCATCT 2934
Oy 2721 -----GCCGAGAGGAGGCTAAGAGAGACACCTCTCATCTGACCAACCTTAATACCTA 2775
Dh 2935 GGTCTGCTGTAGTGTGAAGAAAGTCATGATCCCTTAACAGTATGCTTAATAGTAACAT 2994
Oy 2776 CTGAACCTGTTTGAAGAACCTGCTCTCTCGCAAGAAATGAACAT---CTACCAAGAA 2832
Dh 2995 TTGAAGGTATTTGTAGCTCTTAATCTGGAATAAAGCTAAAGTACATATCATTA 3054
Oy 2833 CTATTGAGACGAATCTCGGAGAACTTCTACGAGAAATGACTGAAGACGACACA 2892
Dh 3055 ACCATTTCTACACAGAGATGAGAAAAATTTATGGAATTTTAAAAAATATATGATACC 3114
Oy 2893 TTCTATTAACGAGAGCTTCACTAATCTGTAATCTAAGCCGATGATATCAACTCTT 2952
Dh 3115 TATTTTAATGATGATATCAAACTGTAATCTAATTAATTAAGTAATTAACAGTTTG 3174
Oy 2953 AACGATATCTAAACGTAAAGAGCTGGAAGAGACATCAATTAAGCTGAAGAAAGCACTG 3012
Dh 3175 ACC---GAACACAAAAAATGCAATTAATGATGAATTAATAAATTAAGATATCTTTA 3231
Oy 3013 CAAGTACGCTGACCTTACAAACAGTACAACTGAACTGAGAGACTCTTGACAAAG 3072
Dh 3232 CAGTTATCATTTGATTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3291
Oy 3073 AAGAAGACAGTGGCAAGTATTAAGTGCAGATCAAGAGAGAGAGAGAGAGAGAGAGAG 3132
Dh 3292 AAAAAAGAACTTGGCCAGACAAATGCAAAATTAATAAATTAATTAATTAATTAATTA 3351
Oy 3133 CTGAAGCAAACTCACTCACTGAAACATCGAAACAGCTACTGACAGACTTCTCACTG 3192
Dh 3352 TTAGAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3411
Oy 3193 TTCTTCAACAG 3252
Dh 3412 TTCTTTAACAATAAAG 3471

Oy 3253 AAGATTTCTTCAAACTACAAAGCCCTCGAAGTATTAATTAATTAATTAATTAATTAAT 3312
Dh 3472 AAAATATTATTAAGACATTAATAAGAGACTTGTAAATATTTATATGTTGATCATCTCA 3531
Oy 3313 CTGAAGCTCTCTCCGAG 3372
Dh 3532 TTAATAAATTTAAGGAGATATCAATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3591
Oy 3373 TTCAAGTCTGCTTAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3432
Dh 3592 TTTAGAGATTAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3651
Oy 3433 AAGCTACCTACCTCTGAG 3492
Dh 3652 AAATATCTTCTTATCAAG 3711
Oy 3493 AAGAACAAGAACTACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3552
Dh 3712 AAAAAATAAATTAATTAAGAGATATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3771
Oy 3553 CTGAATCTTAACAG 3596
Dh 3772 TTAATACTTAAG 3831
Oy 3597 -----CACTGTGTGTCTGAATCTGTGCTCCAGACACTGAGAGAGAGAGAGAG 3641
Dh 3832 CAACAGATTAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCT 3891
Oy 3642 ACCTAAGAA-----GCTGATCTCTCATCTGTCGAGAGAGAGAGAGAGAGAGAGAG 3693
Dh 3892 AAAGAAGAAACAAATATCAACCTTCAAGCTCTTATTAACAGAGAGAGAGAGAGAGAG 3951
Oy 3694 ACATCTCAAGAGCTGAG 3753
Dh 3952 CAATTAACAAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4011
Oy 3754 AGCGAG 3813
Dh 4012 TCCGAGATTAATGACGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4068
Oy 3814 GTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3873
Dh 4069 ACAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4128
Oy 3874 CTGAG 3933
Dh 4129 TTAGCTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4188
Oy 3934 GTCAACGTGAAG 3993
Dh 4189 TTAATTTGAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4248
Oy 3994 TTGAG 4053
Dh 4249 TTAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4308
Oy 4054 CCATTAACAGTTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4113
Dh 4309 TCATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4368
Oy 4114 AAGAGCTCAG 4173
Dh 4369 AAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4428
Oy 4174 CTGAAGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4230
Dh 4429 GTTTTAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4488
Oy 4231 -----CAA 4233
Dh 4489 GAG 4548

LOCUS	DEFINITION	cds, isolate FC27.	5754 bp	mRNA	linear	INV 14-MAR-1994
PFAMMSAA	PFAMMSAA					
LOCUS	PFAMMSAA					
DEFINITION	PFAMMSAA					
cds, isolate FC27.	PFAMMSAA					
VERSION	PFAMMSAA					
KEYWORDS	PFAMMSAA					
SOURCE	PFAMMSAA					
ORGANISM	PFAMMSAA					
REFERENCE	PFAMMSAA					
AUTHORS	PFAMMSAA					
TITLE	PFAMMSAA					
JOURNAL	PFAMMSAA					
MEDLINE	PFAMMSAA					
PUBMED	PFAMMSAA					
FEATURES	PFAMMSAA					
source	PFAMMSAA					

```

gene      /db_xref="taxon:5833"
          1..5754
          /gene="PMMSA"
          <1..5754
          /gene="PMMSA"
          121..5226
          /gene="PMMSA"
          /note="precursor"
          /codon_start=1
          /product="major merozoite surface antigen"
          /protein_id="AA29653.1"
          /db_xref="GI:160413"
          /translation="MKIIIFLCSEFLFIINTOCYTHESYQELAKKLEALDELAVLTGYS"
          LFOEKANLNGSTGTAATVTPGSSGSGGSAVASAYSVSGSGSVASGSGS
          RRTNPDSNDSSTNRYTADLKHRYQNTLYFTIKELAYPELPOLTHMLILSKNVGFKK
          LIDDEELINELLYLNFYIDLAKLADNACANSYCOJLPEKILRANLEADVLRKLQVQ
          RLPDINTIKDYNQKMEDYIKKNTTIANLEIESKRTIDNKNADNEGKRLKLYQAA
          YNLTYNQLOEAHNLIISLEKRIDITLKNKNILLEDIDIKITDENENYTGSKPNN
          LPEKKKEVEGHEEKIKEIARTIKENIDSLTDPLELEYIKREKKKADVPKSDPPK
          KSVQIPVPYNGVITVPLPLDIHNSLAANDKSYGDLMDPDKREKILINDNKK
          RKTINNIKKOIDEKKNINTRQNNKKLLEDYKSKDYELLEKPEKPFNNPDK
          DVVPIFSARTYIVVERKRNKFSNSNSVYNQKALKALSYLEDYSLRQISEKDD
          NHYITLRTGLEADIKLLEIKSSENNKLENNFGLTHSAVASLEVSIVKLOQKVI
          LIKLEIDLKLELFLKNAOLKDSIHVPYIPQPKPEPYLYIVLKEVDLKEEIPKVI
          KMLKBOAVLSSTIOTPLVASSETDEGSHTHLSOGSETEVEEFTVEYGHTEYV
          TWTITLPLKEESAKREKVVENSIEHNSDSOLTKVYKILKDELPTLCTGYCHYK
          LVSNSSMDOKILEVYNLTPREENILKSDPDLPLFNIONTPMYSLYDSNIDLOH
          FLELQKEMITYLKLKEBNHKKLEBOKOITGTSSTSGGNTVTNQSATHSNS
          NQSNASASTNTQNGVAVSSGPAVVEESHDPITLVYSIDNLGIVSLNLRKTVPN
          LTITTEMEKYEENILKNNDYFPNDIKQFVKSYSKYVTGLTEOKNALNDEIKKLK
          TLOSPFLYNNYKLLDLRFLFKKKELGODKKQIKLTLKQLESKLNLSNNPNVLQ
          NSEVFPKKKKEALIEENTLLENTRILKHGKGLVKKYNGE5SPILKTSEVSIOEDN
          YANLEKRALSKIDGKLNLDNLHKKKLSFLSSGLHHLIPLKVIKKNYTGSPSS
          NNKVNENKLSYENFLPEAKVTYVTPPODVPYSPISVRSVSGSGSKETQIPTS
          SLELDOOVLOLNYDEEDSLVVLPTGESENDXYLDQVYTEALSIVTDNLISG
          ENEDYVLYKLQVAGVRSLSKQIENKILTFMNLNLRKRYFLDVLIEDLMQ
          KHLISSENYIIEDESKLNSBOKNTLKSXYIKESVENDIKFAQEGISYEYLAKY
          FDIHSEIKVYIKKEERKPPSPPTTPPATKDBOKRSKRLPLTINLETLYNNLVNN
          IDDLINLAKINDCNVNEKDPDAHRYIKTSLDKAIDDKITDKFNKNDPEAKIKLIND
          TKRMGLKLSLSTGLVONPNTIISKLENGKQDMINLSQHCYKQCPENSGCRHLE
          ERECKCGLANKQSGDKCVENPNPTCNENNGGCDADATCEDEGSSSRKRTYECQYK
          DSYPLPDGIFCFS5NFI5SLFLILMLILV5FI"
          121..177
          /gene="PMMSA"
          178..5223
          /gene="PMMSA"
          /product="major merozoite surface antigen"
          /product="709 c 708 g 1788 t
          BASE COUNT 2539
          ORIGIN 218 bp upstream of HindIII site.

Query Match      26.1%; Score 1290.4; DB 3; Length 5754;
Best Local Similarity 56.4%; Pred. No. 1.le-276;
Matches 2904; Conservative 0; Mismatches 192; Indels 246; Gaps 17;

10 ATGAATCATTTTCTCTCTCTGTCATTCCTGTTTATATACAAATCACTAGGCGG 69
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 ATGAAGATCAATCTCTTTTATGTCATTTCTTTTATATATAAATACAAAGTGTA 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

70 ACCACGATCCCTATCAGAGAGCTGGTTAAGAACTGAAAGCTTTGGAAGATGCCCTCT 129
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 ACACATGAAGATATCAAGAACTTCTCAAAAAAAGTATTAAGAACTTTAGAAATCATATTG 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

130 ACCGATACAGCCTGTCAGAGAGAGATGCTGTAATGAAGGACAGAGTGGAACG 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ACAGGTTTACGTTTATTCCTCAAAAAAAGTATTAAGTAAAGAGAAACAAGTGGAACA 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

190 GCCGTTAAACACGACACACCGGCTCTTAAGGCTGTGAGCTAGCGGTGCTCCGCTGG 249
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 GCTGTAAACATAGTACACCTGGTTCAGGTGCTTACGTTACTTCAAGGTGTTCACTGCT 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

250 TCTGAGGCTCTGGGGGATTCGCTCCGCTCCGCGGACGAGGTGCATACAGTGCCTAGTG 309
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 TCAGTGTCTTCACTGCTCTTC-----AGTGGTTCACAGGTGCTTCACTT 402

```


[illegible]

OY	1366	ATCAACGAAATCAAGAGGAGATCAAAATTTGAGAGACAAATAATTTGACAGTGACAGAA	1425
Db	1483	ATTATATACATTAAAAAAACAAATTTGATTTAGAAGAAAAAAACATTAAATCACACAAAGAA	1542
OY	1426	AGTTACGAAGCCGACCAAAAGCTCAAAAGATATCACTAAAGAGATGAAAGAGCTGCTG	1485
Db	1543	CAAAATAAAAAATTACTTTGAAGTTATGAAAGCTCAAAAAGAGATTTAGMAAATTTACTT	1602
OY	1486	AACGAGATCTATGATTTCCAAATTTCAACAAATTAACATGACCTGACCACTTGAGAAAAATG	1545
Db	1603	GAAAAATTTTATGAAATGAAATTTTAAATTAATTTTGGACAAAGATGTGATGATTAATA	1662
OY	1546	ATGGAAAAAGGTACTTTCCTTCAAGATGGAGAACTGACACCCATTAATCTTTCATCC	1605
Db	1663	TTTCAGTGCAGATATTAACATTAATTAATTTGAAAAACAAAGATATTAATTAATTTTCAATCC	1722
OY	1606	TATGAGATTTCTAGCACTAATCTTTGGAAGCTCACAAAGCTCTTAAGATATGAGAGAC	1665
Db	1723	TCTATATTTCTGTATATTAATTTGTTCAAAAATTTAAAAAGCTCTTTCATCTTGAGAT	1782
OY	1666	TATTCCTCGCGGAACATTTGTTGGAAAAAGAACTAAAGTATTCACAGAATCTCATAGT	1725
Db	1783	TATTCCTTTAAGAAAAAGAAATTTCTGAAAAAGATTTTAACTATTAATTAATCTTGAACCT	1842
OY	1726	AAGATCCAAAAACGATTCGAGAGCGCTTGTGAGACATTTAAGAGAGATGAAGACATGG	1785
Db	1843	GCGCTCCAAAGCTGATATTAATTAATTAATTAATTAACAGAAATTAAGATGAGAAAAACAAAT	1902
OY	1786	TTTGAGAAAGAGATTTACAAAAAGACAAATTAACCGATGAGAGACATCTGAGAGCTCC	1845
Db	1903	CTAGAAAAAAATTTTAAAGACTTAACACATTCAGCAATG---CTTCTTGAGAGTATCT	1959
OY	1846	GATATTTGTTAAAGTCGCAAGTGCGAAGGTGCTCTCATGAAACAGATTTGATGAATCAAG	1905
Db	1960	GATATTTGTTAAATTAACAGATCAAAAAAGTTTATTAATTAATTAATTAATTAAGAGCTTAAGA	2019
OY	1906	AAGACTCACTCATTTCTGAAGAACGTGGAGATTAAAAATATATCAATGTGCCAATAGT	1965
Db	2020	AAGATAGATTAATTTTAAAAAATGCAACAATPAAGATGATTCATCACTACCAACATTTT	2079
OY	1966	TATTAAGCAGGAATTAAGCAAGAACCATCACTCATCTGCTCAAGAAAGAGATTAAGC	2025
Db	2080	TATTAACCACAAAATTAACCAAGAACCATATTTATTTATTTGTTTAAAAAAGAGTATAGT	2139
OY	2026	AAACTGAAAGTGTTCATGCCCAAGTCGAGAGGCTGATCAACGAAGAGAGAAGAAACTT	2085
Db	2140	AAATTTAAAAAGATTTATACCAAAAGTAAAAAGACATGTTAAAGAAAGAACAGCTGTCTTA	2199
OY	2086	AAAACTGAGGACAGTC-----AGATTAATCTCGAGCTCTCCACAGAAAGAGATTAAC	2139
Db	2200	TCAAGTATTTACACAACTTTAGTTGGACGAAGGAAACACTGAAGATGGGGGTCACTCC	2259
OY	2140	GGACAGGCTACACCAAGCCCGGACCAAGGCGGTTCAGCTCTCGAAGGCGATAGGCTG	2199
Db	2260	ACACACACATTTATCCCATATGAGAGAAACGAAGTAAACGAAGAAACAGAAATTAACAGAA	2319
OY	2200	CAAGCTCAAGCAGAGGAGGAAAGGACGACAGCCCTCCAGTGCACAGTCCCGGTTTC----	2255
Db	2320	GAAACAGTAGAGCACACAAACAGGTAAACATTAACATTCACCAAAAGAAAGAAATCAACA	2379
OY	2256	-----AGAGGCTAAAGCTCAAGTGCTTACACCAACAGCTCTGTGATATTAACAGACGAG	2310
Db	2380	CCAAAGAGATTAAGATTTGTTGAAATTTCAATTAAGAAATTAAGATTAAGCAAAATTCACAA	2439
OY	2311	AATGTACGCAAACTGAGCTCCTTGGAAGGCTATGATGCTCTGATATTAACATCTCACTC	2370
Db	2440	GCTTTGACAAAACAGTTTATCTTAAGAAATTAAGATTAATTTTAACTTAATCAATATTA	2499
OY	2371	TGCCACAATATATCTCTCTCTCAGACAGCTATGAACGAGAAAGATTTCTTAACAGTAC	2430
Db	2500	TGTGATTAATATTAATTTAGATCAAACTGTAGATGAGCAAAAAATTTATTAAGAGTATAT	2559
OY	2431	AAGATTAACCAAGAGAGGAGCTAAAGTCTCTCTTGTGATCACTGAGACCTGCTGTTCC	2490

Db	2560	AATCTACTCCAGAAAGAAATAATGAATTAATAATCATGTGATCCATTAGATTATATT	2619
Qy	2491	AATATCAGAAACAACATTCCTCGTTATGTATTTCTATGTGATAGCCTCAACATTTCTC	2550
Db	2620	AATATTCAAAATAACATCACTCGCTAGTGTATTCATTTATGATAGTATGAACTTGATTTA	2679
Qy	2551	TTCTAACTGTTCACTGAGATATATATGAAAGAGATAGCTGTGCACCTGATATTAAC	2610
Db	2680	CACATCTCTTTTGTGATTAATTAACAAGGAATATTTATTTACATTAACCTAA	2739
Qy	2611	GACACAGCAAGATTTAAGACTCTTCGAGG-----	2641
Db	2740	GAGAAATCACAATCAAAAATTTATAGAGACCAAAAACAAATATAGAACATCATCT	2799
Qy	2642	-----AAGCTAAGAGGCTCCACCTGTGTTAAACTCTCTTC	2681
Db	2800	ACATCCAGTCCGTGGAAATTAACACCGTAATATGCTCATCCGACAACAGTAATTC	2859
Qy	2682	CAGCTCATGCAACCACTGTCTCTCACACTCAAGCA-----	2720
Db	2860	CAAAACCAACAAATCAATATGATCCTCTCAACATATCCAAAATGGGTAGCTGATCATCT	2919
Qy	2721	-----GCCCGAAGTGAAGCGCTTAACGAGACCACTCTCACTCGACCAACCTTAATAC	2757
Db	2920	GGTCTCTGTGTAGTTGAAGAAAGTCAATGATCCCTTAACGATATGTCTATTGTATACAT	2979
Qy	2776	CTGAAGCTGTTTGAACATCTGTCTCTCGGCAGAAATTAAGACAT---CTACCAAG	2832
Db	2980	TTGAAAGATATTTGTACTCTTAATACTTGAAATTAACATAAGTAACTAATCATTA	3039
Qy	2853	CTTATTTGGACAGAAATGCTCCGAGACTTACGAGAAATACTGAAGACGACACA	2892
Db	3040	ACCAATTTCTACAAACGATGGAATAATTTATGAAATATTTTAAATAATATGATAC	3099
Qy	2893	TTCTATTAACGAGACTTCACTAATCTCGGAAATCTAAAGCCGATATCAACTCTCT	2952
Db	3100	TATTTTATGTGATATCAACAACATTCGTAAATCTAATTTCAAAAGTAATTAACAGTTG	3159
Qy	2953	AACGATGAATCTAAACGTAAAGAGCTGGAAGAGACATCAATTAAGCTGAAGAGAC	3012
Db	3160	ACC---GAAACACAAAAAATGCAATTAATATGATGAATAATTAATAATTAAGATCTT	3218
Qy	3013	CAACTGAGCTTCGACCTGTCACAAACGATCAAACTGAAGAGACTCTTGACAAAG	3072
Db	3217	CAGTATCATTTGATTTATTAATAATAATAATTAATAATTTGATAGATATTTATATAG	3276
Qy	3073	AAGAAGACAGTCGACAGTAAATAGACGAGATCAACAAGTGTCTGTCTCAAGAGAG	3132
Db	3277	AAAAAGAACTTGCCCAAGCAACAAATGTCAAATTTAAAAAACTTACTTTATTTAAAGACA	3358
Qy	3133	CTTGAAGCAAACTCAACTCACTGAACCAATCCGAAACAGTACTGAGAACTTCAAGT	3192
Db	3337	TTAGATCAAAAATTAATTAATTAATTAACCCCATTAATGTATTAACAAAACCTTTCTGT	3396
Qy	3193	TTCTTCAACAAGAAAGAGAACCCGAGATGCCGAGACAGACACACTCTGAGAACAC	3252
Db	3397	TTCTTTTAAACAAAAAAGAAAGCTGAAATAGCGAAACCTGAAACACATTAAGAAACACA	3456
Qy	3253	AAGATTTCTTCAACACTACAAAGGCTGCTCAATATTTATTAATGGGAGCTTCTCT	3312
Db	3457	AAAATATTATTAGAAACATTTAAAGAGCTTTGTTAAATATTAATTAATGTAATCACTCA	3516
Qy	3313	CTGAAGACTCTCCGAGAGACATCAACAGCGAGATTAATACCGCCGCTCGAGAAC	3372
Db	3517	TTTAAAAACTTTAAGTGAATGATATCAATTTCAACACGAAAGATTAATTTGSCAATTTAGAAAA	3576
Qy	3373	TTCAAGTCTGTCTAAGCTCGAAGCCAAAGCTGAAGACAACTGAACCTGGAAGAAAG	3432
Db	3577	TTTAGAGATTAAGTAAATATAGATGGAACCACTCAATGATTAATTTACATTTAGAGAAAGA	3636
Qy	3433	AAGCTCAAGCTACTCTTAGCGAGATGCACTGATGTGCGGAGCTCAAGGAAGTCA	3492

D6	3637	AAATTACTTTCATTCACTACAGTGATTACAATCATTTTAATTACGAAATTAAAAGAAGTATA	3656
OY	3433	AAGAACAAGAACTCACCGCCATTAACCCAGAGGAATAAATACAGACTGAATPAAGCA	3552
D6	3637	AAAAATATAAATTATACAGGTAAATCTCCAAGGAAAAAATAATAGAAAGTTACGAACT	3736
OY	3553	CTGGAACTCTTACAGAGATGCTCGCTGGAAGAACAGATGTGC-----	3596
D6	3737	TTAAATCTTACGAAATTTTCTCCAGAAAGCAAAGTTACACAGTTGTAACTCCACT	3816
OY	3597	-----CACGTGGTGTCTGAATCTGGCTCCGACACCTGGAGCACTCA	3641
D6	3817	CACACAGATGTAATCATTCATTCCTCATTTATCTGTAAGGGTAAGTAGTTCAGGATCCACA	3876
OY	3642	ACCTAAGAA-----GCCGTGATCTACTCATGTGGAGCCGAGTCCAAATACATTTACC	3693
D6	3877	AAGAAGAAACACAAATACCAACTTCAAGGCTCTTATTTAAACGAATTAACAACAATAGTA	3936
OY	3694	ACATCTCAGAACGTCGACGATGTAGGTCGATGATGATCATATTGTCATCTCGGACAG	3753
D6	3937	CAATTACAAATTATGACGAGAAAGATGATTCCTTAGTTGTATTACCACTTTTGGAGAA	3996
OY	3754	AGCAGAGGAGCTPACGATGTGACTCGGACAGGTGGTCACCGGTAGAGCTGTACTCTTCC	3813
D6	3997	TCCGAAAGTATAGCAATTTTAAAGTACAAAGTAGTAGTACGGAGAAC-- -AATATCTGTC	4053
OY	3814	GTCATTGTATCACTTCTGTCCAAAATCGAGAACGATACGAAGTCTCTATCTGAACCT	3873
D6	4054	ACAATGTATATATCTCTCTAGGATTTGAAAATGAAATATGATATATATTAAACCT	4113
OY	3874	CTGGAGGCGCTTATAGTCTCTCAAGAAACGCTGCAATACGCTGATGTCCTCAT	3933
D6	4114	TTAGCTGGAGATATACAGCTTTAAAAAACCAATTGAAAAAACATTATTACATTTAAT	4173
OY	3934	GTCAACGTGAAGGACATCTCGAACACGCGCTTTAATACAGAGAAATTTTCAAGACGTC	3993
D6	4174	TTAATTTGAACGATATCTTAAATTCACGCTTAAAGAACGAAATATTTCTTAGATGTA	4233
OY	3994	TTGGAGAGCGACTTGAATCCCTTATAAGACTGACTCTCTTACTACGTTGTCAAGAC	4053
D6	4234	TTAGAACTGTATTTAAAGCAATTTAAACATATATCTCTCAAAATGAATATATTAGTAAT	4293
OY	4054	CCATACAAATCTCCCATTAAGAGAAAGAGGATTAATTTCTGTAGTTACACATATTC	4113
D6	4294	TCATTTAAATTATGAATTCAGACAAAAAACACACTTTTAAAAAGTTACAAATATATA	4353
OY	4114	AAGACTCTCATCGACACCGATATCAATTTTCGTAAATGATGTCTGGGTATTACAAGTC	4173
D6	4354	AAAGATCAGTAGAAATGATTAATTAATTTCGACAGAGAGATTAAGTTATATTGAANA	4413
OY	4174	CTGACGCAAAATATCAAGTCTGACTGACTTGAATTTAAAAAGTATATCAAGATAG- --	4230
D6	4414	GTTTTCAGGAATTTAAGATGTGATTTAGATCAATTTTTAAAAAGTTATCAAAAGMAAAG	4473
OY	4231	-----CAA	4233
D6	4474	GAGAGTTCCCATCATCACACACAACACCTCGSTCACAGCAAAAAACAGCAGACAA	4533
OY	4234	GCGCAGATGAAAAATATATGCCCCCTTCGTAAATACATCGAAACCCGTACAAAGCANGTG	4293
D6	4534	AAGAAGAAAGTAGTAAAGTTCCTTCATTTTTRACAAACATTGACACTTATACAAATACCTTA	4593
OY	4294	AAGCACAAAATCGACTCTTCGTAAATTCACCTGGAGGCCAAGSTCTCAACTATACCTTAC	4353
D6	4594	GTTATATAAATTTGACGATTTACTTAAATTAACCTTAAGACAAAGATTAACGATGTAAATGT	4653
OY	4354	GAGAGAGCAATTTGGAAGTTTAAAAATACAGAGCTGAATACCTCAAAAACAATCCAGAC	4413
D6	4654	GAAAAAGATGACGACATGTTAAAAATACAAACTTAGATTTTAAAAACAATTCATGATAC	4713
OY	4414	AAGCTCGAAGTTTTCAGAAAAAATAACAATTCGTGGATATGGCAGACCTGTCTACGAT	4473
D6	4714	AAAATAGATCTTTTAAAAACCTAACGCTTCGAAAGCAATTTAAAAAATTGTAAATATAT	4773

Db 409 GGTCCAAAGTGTCCAAAGTGTACAAAGTCCATCTCTTCAACAAGCTTACCTGCTTCA 468
OY 325 GGGAAACAGTCGAAAGAACCAATCCATCTGACACTCTAGAGATTCGAGCCGCAAGTCTCC 384
Db 469 AATACCTTCATCTGGTGCACGAGCCCTCCAGCTGATGCAAGGATTCAGATCTAAATCTTAC 528
OY 385 GCCGACCTCCAGCAGCCAGTGAAGAACTATCTCTCCTCATCTATCAGAGGATGGAATACCA 444
Db 529 GCTGATTTAAACACAGAGTACGAAATTTACTTGTTCACATTAATGAAGACTCAAAATATCC 588
OY 445 CAGTGTTCGACCTCATATCATATGCTGACACGTGTGATTAACATTTATGCTTCAAA 504
Db 589 GACACTCTTGAATTTAAACCATATGTATTACTTGTGTGATTAATTAATTAATGCTTCAAA 648
OY 505 TATTCGATTTGAGGTTAGCAAGAGATTCATGAACCTCTGACAAAGTTGAATTTCTCTCC 564
Db 649 TATTTAATTTGATGATATGAAAGAAATTAATGAATTTATTAATTAATTAATTTATTTT 708
OY 565 GACTTGTCTAAGGGCCAACTGATGACGTTTGCCGCAATGACTATTTGCAAAATTCATTC 624
Db 709 GATTTATTAAGAGCAAAATTTAATGATGTATGTGTAATGATTAATTTGCAAAATCTTTC 768
OY 625 AATTTGAATATCAGAGCCAGCAAGTGTGACGTATTTGAAGAAGTGTCTTCGATATCCG 684
Db 769 AATCTTTAAATTCGTGCAATGATTAATTAAGCTTAATAAACTGTTTGGATTTAGA 828
OY 685 AAGCCTCTCGACACATCTCAAGACATGTGGAAAGATGGAAGATTTATTAATAAAGAT 744
Db 829 AATCCATTTAGCAATTTTAAAGATATGTAGGAAATGGAAGATTTACATTTAAATAAAT 888
OY 745 AAGAGACCATGAGACATTTAAACGAGTGTGAGAAATTCCAAAAGCCATAGCAAA 804
Db 889 AAAACACCATAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 948
OY 805 AATTAAGATTCACACAGAGAGAGAAAGAAAGATGTGACAGGCCAGTGCAGCTG 864
Db 949 AATTAAGATTCAGATTAATTAAGAGAGAGAAAGAAATTTATACCAAGCTCAATTTGATCT 1008
OY 865 TCCATCTATTAACAAACAGCTTGAAGAGCCCATACCTCATGACGCTACTGAGAGAGCC 924
Db 1009 TCTATTTCAATTAACAAATTAAGAGAGACATTAATTTAATAAGCTTTAGAGAAACGT 1068
OY 925 ATAGACACCTCATAGAGAAAGATTAATTAACAAAGATCTGTCAGCAAGATTAATGAAT 984
Db 1069 ATTGAACCTTTAAAAAATAAAGATTAAGAAATTAATTAATTAATTAATTAATTAATTA 1128
OY 985 AAGAACTCTCCGCGCAACTCTGGAGACACCCCTAACAAGCTGTGAGCAAGACAA 1044
Db 1129 AAAAATCCCCACGCGCCCAATTTCTGGAATACACCAATATCTCTTGTATAGAACAA 1188
OY 1045 AAGATTAAGAGACAGAGAAAGATCAAGAGATGCGCAAAACATTAAGTTCAACAT 1104
Db 1189 AAAATTCGAGGACACGAAAGAAATTAAGAAATTTGCAAAATCTTTAAATTTAACTT 1248
OY 1105 GATTCCTCTTATCTGATCCCTTGGAGTGTGAGTACTACTGTGAGAGAGAAATTAAGAT 1164
Db 1249 GATAGTTTATTTATCTGATCCACTGATTAATTAATTAATTAATTAATTAATTAATTA 1308
OY 1165 ATAGACATCTCCGCAAAAGTCA-----GACAAAGAAATCAACCGAATCTTAATGA 1215
Db 1309 GTTGATGTAACACCTAAATACAGAAATCTGAGAAATCTGTTCAATTAACAAAGTTCC 1368
OY 1216 TATCCCAATGTGTGAGTACCTCTGTCTTATTAACGATTAACAAAGCTCT----- 1268
Db 1369 TATCCAAATGTGTGTATATCTTATACCACTACTGATATTCATTAATTAATTAAGTGA 1428
OY 1269 --CAAGAGCTCAATAGCTTGGTACTGTGATTAACCCCTGATTAATAGAGAAACCC 1326
Db 1429 GATTAAGATTAATAATTCATATGATTTAATGAATCTTCATTAAGAGAAATTAAT 1488
OY 1327 TCTTAAGATTAATTAACAGAGATGAGAGAAAGATTTATCAACGAATTAAGAGAG 1386
Db 1489 GAAAAAATTTATTTACAGATTAATTAAGAGAAAGAAATTTATCTATTAAATTAATAA 1548

OY 1387 ATCAAAATTTAGAGAGAGAAATTTAGAGTGAAGAAAGATTTAGAGACCGACAGCAAA 1446
Db 1549 ATTGATTTTGAAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1608
OY 1447 AGCTTAACGATTAATCAATTAAGAGATTAAGAAAGTGTGAGAGATCTATGATTTCAAA 1506
Db 1609 GATTTATGAAGAAATCAAAAGAAAGATTAATTAATTAATTAATTAATTAATTAATTA 1668
OY 1507 TTTCAAAATTAATGAGTGAAGAAAGTGTGAGAAAGTGTGAGAAAGTGTGAGAAAGTGT 1566
Db 1669 TTTAATTAATTAATTTGACAAAGATGCTGATTAATTAATTAATTAATTAATTAATTA 1728
OY 1567 AAAAGTGAAGAACTGACACACCATTAATTAATTAATTAATTAATTAATTAATTAATTA 1626
Db 1729 AATGTTGAAGAAAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1788
OY 1627 GTTGAGACCTACCAAGCTCTTAAGTATGAGAGATTAATTTCTGCGGAACATTTCT 1686
Db 1789 GTTCAAAATTTAAAAAGGCTCTTCTCATATCTGGAAGATTAATTTTAAGAAAGAAAT 1848
OY 1687 GTTGAGAGAAAGTAAAGATTAACAGAAATCTCATTAAGTAAAGTGAAGAAAGATCGAG 1746
Db 1849 TCTGAAAAAGATTTTATATCATTTATTAATTAATTAATTAATTAATTAATTAATTA 1908
OY 1747 ACGCTTTGAGAACATTAAGAGATTAAGAGATTAAGAGATTTGAGAGAAATTAACAAA 1806
Db 1909 AATTTAAACAGAAAGAAATTAAGAGATTAAGAGATTAAGAGATTTGAGAAAGAAATTT 1968
OY 1807 GAGGAAATTAACCCAGATTAAGAGAAAGTGTGAGAGTGTGAGAGATTTGAGAGATTT 1866
Db 1969 CTATACATTAACAGATTAAGAGATTTGAGAGATTTGAGAGATTTGAGAGATTTGAG 2025
OY 1867 CAGAGAGTCTCTCATTAAGAGATTAAGAGATTTGAGAGATTTGAGAGATTTGAGAG 1926
Db 2026 CAAAAAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2085
OY 1927 AAGCTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1986
Db 2086 AATGCAAACTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2145
OY 1987 GAACCTACTACTCTCATCTGAGTCAAGAGAGATTAAGAGATTTGAGAGATTTGAGAG 2046
Db 2146 GAACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2205
OY 2047 AAGTGTGAGAGCTGTATCAAGAGAGAAAGAAAGATTAATTAATTAATTAATTAATTA 2102
Db 2206 AAGTAAAGACATGTTAAAGAGAAAGAAAGCTGTCTTATCAAGTATTAACAACTTTA 2265
OY 2103 --AGATTAACCTCGAGCTTCCAGAGAGAGATTAAGAGAGCTACACCAAGCC 2160
Db 2266 GTTGACAGCAAGGAAACAACTGAGAGTGGGGTCACTCCACACACATTTATCCAAATCA 2325
OY 2161 GCACACAGGCGGCTTCACTCTGAGAGGCGATAGGCTGACACCTCAACACAAAGCAG 2220
Db 2326 GAGAGAAACAGAGTAAACAGAGAGAAACAGAGAAAGAAAGATTAAGAGACACACAA 2385
OY 2221 AAGCAGC-----ACAGCTCCAGTGCAGTGCCTGTTCCAGAGGCTAAAGCTCA 2271
Db 2386 ACGTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2445
OY 2272 GTTCCATACACACACCTCTCTGATTAACAGAGAGAGATTAAGAGATTTGAGAAATG 2331
Db 2446 GAAATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2505
OY 2332 CTGAGAGAGCTATGAGTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2391
Db 2506 CTAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2565
OY 2392 TCTCAGACAGTATGAAGAGAGATTTTAAACAGTATCAAGATTAACAGAGAGAG 2451
Db 2566 TCAACTCTAGTATGAGACCAAAATTAATTAAGATTAATTAATTAATTAATTAATTA 2625

OY	2452	AGGAAGCTGCTCTGTGTGATCCGAGACGCTGCTCAATATCCAGAACACTTCC	2511
OY	2453		
Db	2626	AATGAAATTTAAAATCATCTGTGATCCATTAGATTATTTATTTAAATATTTCAAAATTAACATTA	2685
OY	2512	GTTATGATTTCTATGTTCCGATAGCTTCACAAATTTCTCTCACTGTGATGGAGTA	2571
Db	2686		
OY	2512	TATGAGAGGAGATGTCCTCAACTGTATTAACCTCAAGAACAGACAGATTAAAC	2631
Db	2746	TATCAAAAGGAAATGATTTATTTATTAACATTAACCTAAAGAGAAATACATCAAAA	2805
OY	2632	CTTCTGAGGAGACTAAGAGCTC-----	2657
Db	2806	TTATTTAGAGGACAAAATAAATTAACGTGACATCATCTCACTCTGGAAATTA	2865
OY	2658	-----CACCTGTGTAAACCTCTCTTCACCTCCATGCAACCACTGTC	2702
Db	2936	TCTCTACCAATACCCAAATGCTGATGCTATCTGTCGCTGCTGATGTTAGAA	2985
OY	2737	AACGACGACACTCTCACTGACCAACTTAATTACTCAAACTGTTTGAGACATC	2786
Db	2986	AGTCATGATCCCTTTAACAAGATTTGCTATTTAGTAAGATTAAAGATATTTGTTAGTTC	3045
OY	2797	CTGCTCTCGCGCAAGATTAAAGACAT---CTACCAAGACTATTGACAGAAATCGTC	2853
Db	3046	TTAAATCTTGGAAATTAACCTTAAGTACCTATCTCACTTAACATTTCTACACAGAGATG	3105
OY	2854	GAGAACTTCTACGAGAGATACTGAAAGACAGGACACATTCATTAAGACAGAGCTTCACT	2913
Db	3106	GAAAAATTTATGAGATATTTTAAAAAATAAGATACCTATTTAATGATGATATACAA	3155
OY	2914	AACCTTCTGAATTTAAAGCCGATGATATCACTCTTTAACGATGAATCTTAACGTAAG	2973
Db	3166	CAATTCGTAAATTTCTAATTTCAAAAGTAATACAGGTTTGACC---GAAACACAAAAAAT	3222
OY	2974	AAGCTGGAAGGACATCATTAAGCTGAAGAAGACCTGCACTGAGCTGAGCACTGAC	3033
Db	3223	GCATTTAATGTGAATTTAAAAAATTTAAAGATACTTTCACATTTATGATTTATAT	3282
OY	3034	AACAAGTACAACACTGAACCTGAGAGACTCTTCGACAGACAGACAGAGCTGGCAAGAT	3093
Db	3283	AATTAATATTAATTTAAATATAGATGATTTATTTAATAGAAAAAAGAACTTGGCCAGAC	3342
OY	3094	AAGTTCAGATCAAGAACTTGACTCTGCTCAAGAGACAGCTTGAAAGCAAACTCACTCA	3153
Db	3343	AAATTCGAAATTTAAAAAACTTACTTTATTTAAAAAGAACATTTAGATCAAAATTTGAATTC	3402
OY	3154	CTGAACAATCCGAACCGCTACGACGAACCTCTCAGTGTCTTCAACAGAGAGAGAA	3213
Db	3403	CTTATATACCCACATTAATGTATTACAAACTTTCTGTTTCTTTAAACAAAAAAGAA	3462
OY	3214	GCCAGATCCCGCAGAGACGAGACACTCTGAGAGACACCAAGATTTCTTCGACACATAC	3273
Db	3463	GCTGAATATAGCAGAACTGAAACACATTTAGAAACACAAAAATTTATTGAAACATTAT	3522
OY	3274	AAAGGCTCTGCAATTAATTAATTAATGCGAGCTCTCTCTCTGAAGACCTCTCCGAGAG	3333
OY	3334	AGCATCCAGACCGAGATTACTACGCCACGCTCGAGAACTTCAAGTCTCTGTAAAGTCTC	3393
Db	3583	TCAATTTAAACAGAGATTAATTAATGCAATTTTAGAAAAATTTTAGATTAATAATA	3642
OY	3394	GAAAGCAAGCTGAAGACATCTGAACTCGAGAGAGAGAAAGACCTGAGTACTCTGTAGC	3453
Db	3643	GATGGAATAACTCAATGATTAATTTACATTTTAGGAAAAAATAATATCTTCTTATCAAGT	3702
OY	3454	GGACTGATCACTGATCGCGAGCTCAAGGAAGTCAATTTAAGAACAGAACTACACGGC	3513

Db	3703	GGATTACATCTTTAATTACTGGAATTTAAAGAAAGTAATATAAAATATATACAGTT	3762
Qy	3514	AATAGCCCAACCGGAATTAATACAGACGTAAATAAGCACTGGAAATCTTACAAGAATTC	3573
Db	3763	AATTCTCCAACTGGAATAATTAAGAAAGTTACGAAGCTTTAAATCTTACAGAAATTTT	3822
Qy	3574	CTGCGTGAAGAACAGATGTGCGC-----CACTGT	3602
Db	3823	CTCCCGAAGCAAAAGTTACAAAGTTGTAACTCCACCTCAACACAGATGTAACTCACTCT	3882
Qy	3603	GGTGTCTGAATCTGGCTCCGACACACTGGAGCAGTCTCAACCTAAGAA-----GGCT	3654
Db	3883	CCATTATCTGTAAAGGGTAAGTGGTAGTTCAGGATCCAAAAAGAAAGAACACAAATACCA	3942
Qy	3655	GCATCTACTCTATGTGCGGAGCGAGCTCCAAATACATTAACCAATCTTCAGACGTCCAGCAT	3714
Db	3943	ACTTCAGGCTCTTTATTTAACGAAATTTACAAACAGTAGTAGTAATTTACAAATTTATACGAA	4002
Qy	3715	GAGTCGATGACGTCATCATTTGTGCTATCTCTGCGGAGAGCGAGAGAGGACTAGCATGAC	3774
Db	4003	GAAGATGATCTCTTAGTTGTATTACCATTTTGGAGAAATCCGAAGATTAATGACGAATAT	4062
Qy	3775	CTCGGCCAGGTGTGTCACCGGTGAGGCTGTACCTCTCCGTGATGTATACATTTGTGCC	3834
Db	4063	TTAGTCTCAGTAGTAAGTGTGGAGAGCAAAATCTCG-----TCACAATGGATTAATCTCTCA	4119
Qy	3835	AAAATCGAAGCAAGATTCGAAAGTCTTATCTGAAACCTCTGGACGGCTGTATGTCCT	3894
Db	4120	GGATTTTGAATAATGAAATTTGATGTTATATATTTAAAACTTTTACCTGAGCTATATGAAC	4179
Qy	3895	CTCAAGAAACAGCTGGAGAAATTAACGTGTAGCTTCCTCAATGTCAAGCTGAAGCAATTCGTG	3954
Db	4180	TTAAAAAACAATTTGAAAAACAATTTTACATTTAATTTAAATTTGAACAGATATCTTA	4239
Qy	3955	AACAGCGCTTTAATTAAGAGAGAAATTTTCAGAACGCTTTGGAGAGCACTTGATTGCC	4014
Db	4240	AATTCCGCTTTTAAAGAACGAAAAATATTTCTTGATGTATTTAGAAATCTGATTTAATGCA	4299
Qy	4015	TATTAAGACCTGACCTCCTCACTACGTGTGTCMAAGGCCCTTACAAAGTTCCTCAATAA	4074
Db	4300	TTTAAACATATATCTCTTAAATGATATCATTATTGAAGATTTCACTTTAAATTTATGATTTCA	4359
Qy	4075	GAGAAGAGGATTAATTTCTGTCTAGTTTACACTATATCAAGAGCTCCATCGACACCGAT	4134
Db	4360	GAACAAAAAACAACCTTTTAAAAAGTTAAATATATTAAGAAATCAAGTACGAAAAATGAT	4419
Qy	4135	ATCAATTTGCTAATGATGTGCTGGGTATTTACAGATCCTGACGAAAAATACAAGTCT	4194
Db	4420	ATTTAAATTTGCACAGGAAGGTATTAAGTTATTATGAAGAAAGTTTGTGCGAAATATAMGAT	4479
Qy	4195	GACCTTGACCTATTTAAAGGTATATCAACGATAAG-----	4230
Db	4480	GATTTAGATCAATTTAAAAAAGTTATCAAGAAAGAAAGAGAGAGTCCCATCATCACCA	4539
Qy	4231	-----CAAGCGAGAAATGAAAAATATCTG	4254
Db	4540	CCAACACACCTCCGTACCAACAAAAACAGAACGAAACAAGAAAGAAAGTAAGTTCCTT	4599
Qy	4255	CCCTTCTGAATTAACATGGAACCCGTGTACAGACAGTGAACGACAAAAATCGACTCTTC	4314
Db	4600	CCATTTTAAACAAACATTTGAGACCTTATCAATTAACCTTAGTTAATTAATTTGACGATTAAC	4659
Qy	4315	GTAATTCACCGGAGGCGCAAGTCTCAACTATACTTACGAGAAAGCAATGTGGAATTT	4374
Db	4660	TTAATTAACCTTAAAGGCAAGGATTTAACGATGTGAATGTTGAAAAAAGATGAAGCACATCTTT	4719
Qy	4375	AAAAATCAAGAGAGCTACTCTCAAAACAAATCCAAACAGACAAACCTGGCAGATTTTCAGAAA	4434
Db	4720	AAATTAACATACTTATGTGATTTTAAAAACATTTGATGACAAATATAGATCTTTTAAAAAC	4779
Qy	4435	AATAACAATTTCTGTGGAATTTGACAGCTGTGTACCGATTTATTAACACCAACATCTCTG	4494

Db 2815 CAAGAACCAATCAATGATCTCTTACCAATACCCAAATGGTGTAGCTGATCATCT 2874
Oy 2721 -----GCCGAAGTGAAGCCCTAAGCAGACACCTCTCACTCGACCAACCTTAATATCTCA 2775
Db 2875 GGTCCGCTGTAGTATGAAGAAAGTCAATGATCCCTTAACGATATGCTATTAAGTAAACGAT 2934
Oy 2776 CTGAACCTGTGGAGAACATCTCTCTCGGCAAGAAATAGAACAT---CTACCAAGAA 2832
Db 2935 TTGAAGGATTTGTGTAGTCTTAATCTGGAAATTAACCTAACTACCTAATTCATTA 2994
Oy 2833 CTATATGACAGAAATCGTCCGAGACCTCTACGAGAAATCTGAAAGACACCGACACA 2892
Db 2995 ACCATTTCTACACACAGATGGAAAAATTTATAGAAATATTTTAAAAATATATGATAC 3054
Oy 2893 TTCTATACAGAGACTTCACTACTTCTGTAATCTAAGCCGATATATCACTCTT 2952
Db 3055 TATTTTAATGATGATTCAAACAATTCGTAAATCTAATTCAAACAATTAATTCAGSTTGG 3114
Oy 2953 AACGATGAATCTAAGCTAAGAACTGGAAGAGACATCAATAGCTGAAGAAACACTG 3012
Db 3115 ACCGAACACAAAA---AATGCATTAATGATGAATTAATAAATTAAGATACTTTA 3171
Oy 3013 CAACGTAGCTGACCTGTACACAGTACAACTGAACCTGAGAGACTCTCGACAG 3072
Db 3172 CAGTTATCATTTGATTTAG 3231
Oy 3073 AAGAACACAGTCGCAAGTATATAGATCAGATCAGATCAGAAAGTTGACTCTGCTCAAGAGAC 3132
Db 3232 AAAAAAGAACTTGGCCACAGCAAAATGCAAAATTAATAAATCTTATTAATAAAGAACAA 3291
Oy 3133 CTTGAAAGCAACTCACTCACTGAAACAATCGAAACAGTACTGACAGAACTTCTACAG 3192
Db 3292 TTAGAATCAAAATGATTAATTCATTAATTAACCCATATATATTAACAAATCTTCTGTT 3351
Oy 3193 TTCTTCAAGAAAGAAAGAGCGAGATCGCGAGACAGAACTCTGGAGAACAC 3252
Db 3352 TTCTTTAACAAAAAAGAGAGCTGAATATGAGAAATGTAAGAACTAGTAAGAAACACA 3411
Oy 3253 AAGATTTCTCTCAACACTACAAAGAGCTCTGCAAGTATATATATATATATATATATATAT 3312
Db 3412 AAATATATATGAAACATTAATAGAGACTTGTAAATATATATATATATATATATATATATAT 3471
Oy 3313 CTGAGACCTCTCCGAGAGAGACATCCAGACCGAGAGATACCTGACCCAGCTCGAGAAC 3372
Db 3472 TTAATAAACTTTAAGTAAAGTATCAATTCAAACAGAAATATATATCCAAATTTAGAAAA 3531
Oy 3373 TTCAGAGCTCTGTAGAGTCGAAGGCAAGCTGAAGCAAACTGAACCTGAGAGAGAG 3432
Db 3532 TTTAGAGTATTAAGTAAATAGATGAAGAACTCAATGATATATTAATTTAGAGAAAGAAA 3591
Oy 3433 AAGCTCAGCTACTCTAGCGGAGCTGATCACCCTGAGCGGAGCTCAAGAGATCAT 3492
Db 3592 AAATATATCTTTCTTATCAAGTGAATACATCACTAATTAATTAATTAATAAAGAAATATA 3651
Oy 3493 AAGACAAAGACTACACCGGCAATAGCCCAAGGAGAAATATACAGACGTGAATTAACGCA 3552
Db 3652 AAAAAATAAAAATATATACAGTAAATCTCCAAAGTAAATATATATATATATATATATATAT 3711
Oy 3553 CTGAGATCTTACAGAAAGTTCTGCTGAGAAAGAAAGATGTGCG----- 3596
Db 3712 TTTAAATCTTAGCAAAATTTCTCCAGAGCAAAAGTTTACAAAGATTTGTAACCTCACT 3771
Oy 3597 -----CACTGTGTGTCTGAATCTGCTCGACACACTGAGGAGCTCA 3641
Db 3772 CAACGAGATGTAATCTCATCTCATTTATCTGTAAAGGTAAAGTGTGATCAGATTCACACA 3831
Oy 3642 ACCTAAGAA-----GCCTGATCTACTCATGTGGAAGCGAGTCCAAATACAAATTAAC 3693
Db 3832 AAGAAGAAACAACTATACCACTCAGGCTCTTTATTAACAGAAATTAACAAAGATAGTA 3891
Oy 3694 ACATCTCAACAGCTGAGCATGAGTGTGATGAGTCACTATTTGCTATCTCTGCGGAG 3753
Db 3892 CAATTAACAAATTAATGACGAGAGAGATCTCTTATGTTGATTTACCATTTTGGAGAA 3951

Oy 3754 AGCAGAGAGACTACGATGATGACCTCGGCAGGTGTCACCGGTGAGGCTGTCACTCTCC 3813
Db 3952 TCCGAGATATATGCAAGATATTAGATCAAGTGTAGTACGTGAGAGAACAAAT---ATCTGTC 4008
Oy 3814 GTGATTTGATATACCTTTGTGCCAAATTCGAGAACGAATAGAGAGTGTCTATCTGAACCT 3873
Db 4009 ACAATGATTAATATCTCTCAGGATTTGAAAAATATAGTGTATATATTTAAACCT 4068
Oy 3874 CTGCGAGCGCTATATAGTCTCTCAAGAAACACCTGAGAAATACGTGATGACCTCAAT 3933
Db 4069 TTAGCTGAGATATATGAAAGCTTAATAAAGAAATGGAAGAAATTTTACATTTAAT 4128
Oy 3934 GTCAACGTGAAGACATCTGGAACAGCCGCTTTAATTAAGAGAAATTTCAAGAACTG 3993
Db 4129 TTAATTTGAACGATATCTTAATTAATTCAGCTCTTAAGAAAGAAATATTTCTTAGATGTA 4188
Oy 3994 TTGAGAGGAGCTGTGATTCCTTAAGAACTGACCTCTCACTACGTGTCAAGAGAC 4053
Db 4189 TTAGAATCTGATTTAATGCAATTTAACAATATATCTCAAAATGAATCAATTTGAAGAT 4248
Oy 4054 CCATACAGTCTCTCAATTAAGAGAGAGGATAAATTTCTGTCTAGTTACAATATATC 4113
Db 4249 TCATTTAATTAATTAATGATTCAGACAAAAAACAACACTTTTAAGAGTTACAATATATA 4308
Oy 4114 AAGACTCCATCGACACCGATATCAATTTGCTAATGATGTCTGGGTATTAACAAGATC 4173
Db 4309 AAGAAATCAGTAAAGAAATATATTAATTTGACAGAGAGATTAATGATTAATGAAAG 4368
Oy 4174 CTGAGGAAAAATCAAGTCTGACCTTGAATTAAGAAAGTATATCAACGTAAG--- 4230
Db 4369 GTTTTAGCGAAATATATAGGATGATTTAGATTCATTAATAAAGTTATCAAGAGAAAG 4428
Oy 4231 -----CAA 4233
Db 4429 GAGAACTTCCATCATCACCAACACACCTCGCTCACACCAAAAAACAGAACAA 4488
Oy 4234 GCGGAGATGAAAAATATCTGCCCTTCTGATATACATGAAACCTGTACAGACAGT 4293
Db 4489 AAGAAAGAAAGTAAAGTCTCTTCCATTTTAACAAATGAGAGCTTTATACAAATTACTTA 4548
Oy 4294 AAGCAAAATGAGACCTCTGTATTCACCTGAGAGCCAGCTCTCAACTATCTTAC 4353
Db 4549 GTTAATTAATTAATGACATTAATTAATTAATTAAGGCAAAAGTTAATGATTTATGTT 4608
Oy 4354 GAGAAAGCAATGAGAGTAAATTAATTAAGAGACCTGACTACCAAAATCAAGAC 4413
Db 4609 GAAAGAGATGAAACACATGTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 4668
Oy 4414 AAGCTGGAGATTTCAAGAAATTAACAATTTGCTGGAATTCAGACCTGTCTACCGAT 4473
Db 4669 AAATATGATCTTTTAAACCAATTAAGACCTTGAATGCAATTTAAAAATGATTAATGAT 4728
Oy 4474 TATTAACCAACAATCTGCTGACCAAGTTCTGACAGGAGTGTGCTGCAAAACCTC 4533
Db 4729 GATACGAAAAAGATATGCTGCGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 4785
Oy 4534 GCCAAACAGTGTGAGCAATCTGCTGAGCGCAACCTGAGGAGCTGTAACATCTCC 4593
Db 4786 CTTAATACATATATATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4845
Oy 4594 CAGCAACCAATGCTGAGAAACAGTCCCCAGAAATAGCGGCTGTTTACAGCACTGAGAC 4653
Db 4846 CAACACCAATGCTGTAATAAACAATCTCAGAAATTTGAGATGTTTCAAGACTTTAGAT 4905
Oy 4654 GAGCGGAGAGAGTGAAGTCTCTGCAACTTCAACAAGAGAGATATAGTCCGTTGAG 4713
Db 4906 GAAAGGAAAGAAATGTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4965
Oy 4714 AACCCAACCTCTGCAATGAAACAATGAGCGGTGTGACCCGATGCTTAATGACAC 4773
Db 4966 AATTCAAATCTACTTGTATACGAAATTAATGTGTGATGTGATGACATGCGCAATGTAC 5025

[illegible][illegible]

Db	5368	TATCACCCTTTTCGATGATTTTCCTGGACGTTCCCTCCTACCTTCCTAGGAATATCATCTTA	5427
0y	4894	CTGATCTTCATGTCGATCTCCTGACAGCTTCATCTATATGATCATG	4939
Db	5428	TTAATACATCATGTTAATATATATACAGTTTCATTTAAAAAAGTAGG	5473
RESULT 17			
LOCUS	AF480451	1968 bp	mRNA
DEFINITION	Plasmodium falciparum FCB-1 merozoite surface protein 1 (MSP-1)		linear
ACCESSION	AF480451		INV 14-MAR-2002
VERSION	AF480451		
KEYWORDS	AF480451.1	GI:19423278	
SOURCE	Plasmodium falciparum.		
ORGANISM	Plasmodium falciparum.		
REFERENCE	Plasmodium falciparum.		
AUTHORS	Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
TITLE	1 (bases 1 to 1968)		
JOURNAL	Li, X.R., Liu, S.C., Chisht, A.H. and Oh, S.S.		
FEATURES	Submitted (04-FEB-2002) Section of Hematology/Oncology Research, St. Elizabeth's Medical Center, Tufts University School of Medicine, 736 Cambridge Street, Boston, MA 02135, USA		
source	location/Qualifiers		
	1..1968		
	/organism="Plasmodium falciparum"		
	/isolate="FCB-1"		
	/db_xref="taxon:5833"		
	/country="Colombia"		
	1..1968		
	/gene="MSP-1"		
	<1..>1968		
	/gene="MSP-1"		
	/codon_start=1		
	/product="merozoite surface protein 1"		
	/protein_id="AAI8845.1"		
	/db_xref="GI:19423279"		
	/translation="VTHEHYOELVKLLEDAVLVGLSLFQREKYNLVNGSTGATVAT		
	TSPTGSKSVASGSGSVASGSGSVASGSGSVASGSGSGRTNSDSD		
	AKSTADLKHRYRNLITIKELRYQLDRLTHMLTCDNLHGRFLYLDGIEINELLY		
	KLNVFDLIRAKLNDVCAANDCOIPENLKI RANELDVAKLIVGYR PLNDIKNVGR		
	MEDYIKRKKTIENINELIESEKRTIDKNKATSEERKLIYQYRLSYNQLDEEA		
	HNLSVLEKRIIDLEKKNENIKELDKIDINENPNPANGPTNLIDRNNKIESEKE		
	IKELAKTIKFNIDSLFDPLIEYLYREKNNIDISAKVEKTESENPNGVYPL		
	SYNDINNALNLSFGLINPEPYTKEPSKNIYDNEKKRINIEIKRIEKKIES		
	DKKYEYDSKINDITKEYEKLMEIYDSKNNIDITNPEKKMGKSYKVEKLTTHH		
	NTPRSYENSKINDLEKRLKALKRYMEDYSIRNIVAREKLIYKKNLSIKENLETVENI		
	KTRDEQLPEKRIKIDENKPKDKILLEVSDIYAVQYQKVLNKKIDELKKTOLLNVEL		
	KHNHHPVSPYQEKQEPYLVILVLEKDKLK"		
BASE COUNT	894 a	220 c	277 g 577 t
ORIGIN			
Query Match	23.7%	Score 1169.6;	DB 3; Length 1968;
Best Local Similarity	74.6%	Pred. No.3.6e-246;	
Matches 1469; Conservative	0;	Mismatches 499;	Indels 0; Gaps 0;
0y	67	GTGACCCACGATCTCTATCAGAGCGTGGTTAAGAACTGGAGCTTTTGAAGATCCCGTC	126
Db	1	GTAACACATGAAGTTATCAACAAGTTGTGCAAAAAAGTAAAGAGCTTTAGAACATGCACGA	60
0y	127	CTTACCGGATACAGCGCTGTTCAGAGAGAGAGATGCTGATGAATGAAGGACGAGTGCC	186
Db	61	TTGACAGGTTATAGTTTATTTCAAAAGGAAAAAATGATTAATAAGGAAGAACAGTGGA	120
0y	187	ACGGCGGTATACACACAGACACACCGGCTTAAAGGCTCTGTGCGCTAAGCGGTGGCTCCGCT	246
Db	121	ACAGCTGTTACACATGACATCCTGTTTAAAGGGGTTCAGTTGCTTCAGGTGCTTCAGGT	180

Oy	247	GGGTCGTGGCCCTCTGGGGGTTCCGTCGGCTCGCGGGGAGCGTGGCATCGAGTGGGCTCA	306
Db	181	GGCTCACTCTCTTCAGGTGGCTCAGTTCCTCAGGGGGCTCAGTTCCTCAGGTGGCTCA	240
Oy	307	GTGCGCAAGCGCGGGTTCGCGGAACACTCGAAGAACCAATCCATCTGCACACTTAGCGAT	366
Db	241	GTTCCTCTCAGTGGTGGTTCAGGTATTCACAGCGTACAAATCCCTTCAGATTAATTCAGGTAT	300
Oy	367	TCGAGCGCCAAGTCTCAGCCGACCTCAAGCACACCGAGTGAAGAACTATCTCCTCACTATC	426
Db	301	TCGATGCTCAAACTTACGCTGATTTAAACACAGAGTGCAGAAATTAAGTTGTAACATATC	360
Oy	427	AAGAGGTGAAGTACCCACAGTGTTCGACGCTCACTAATCAATCTCAGACCTGGTGAT	486
Db	361	AAGAACTCAATAATCTCTCACTCTTTGATTTAACTAATCATATGTTAATCTTTGGTGTAT	420
Oy	487	AACATTCATGGCTTCAAATTCCTGATTGACGGTTACGAGAGATCAATGAATCCCTGTAC	546
Db	421	AATATCATGGTTTCAAAATTTTAATTTGATGGATGAGAAATTAATGAATTAATTAATAT	480
Oy	547	AAGTTGAATTTCTACTTCGACTTGCCTAAGGGCCAAATCAATGACGTTTGGCCCAATAC	606
Db	481	AAATTAACCTTTATTTTGTATTTATTAAGACAAATTAAGATGAATGTATGTCTAATATAT	540
Oy	607	TATTCGAAATTCATTCATTTTGAAGATCAGAGCCACAGATTTGGAGATTTGAAAGAG	666
Db	541	TATTCGAAATTCCTTTCATCTTTAAATTCGTGCAAAATGAATTAAGAGCTACTTAAAAA	600
Oy	667	TTTGCTCTTGGATATTCGCAAGCCTCTCGACACATCATAGGACAAATGTGGGAAAGTGGAA	726
Db	601	CTTCTGTTCGGATATATGAAAAACCATTAGACAAATTTAAAGATTAATGTAGGAAAAATGGAA	660
Oy	727	GATTATTTTAAAAAGATTAGAAGACCATCGAAGAACATTAAAGACTCATCGAAGAAATCC	786
Db	661	GATTATCTTAAAAAAATTAAAAAACCATTAGAAAAATTAATGAATTAATTGAAAGAAAGT	720
Oy	787	AAAAAGACCATAGACAAAAATTAAGAAATGCACCAAGAGGAGGAAAAAGAAAGTTGTAC	846
Db	721	AAGAAACAATTTGATTAATAATTAAGAAATGCACATTAAGAGAGGAAAAAAATTAATATAC	780
Oy	847	CAGGCCCCAGTACGACCTGCTCATCTATTAACAAANAGCTTGAAGAGGCCATTAAGCTCATC	906
Db	781	CAACCTCAATATGATCTCTTCTTATTTACAAATTAACAAATTAAGAAAGACATATTAATTAAT	840
Oy	907	AGCCTACGTGGAAGCCCATAGACACCCTCAAGAAATGAAGAAATATCAAGAAAGACTGCTC	966
Db	841	AGCGTTTAAAGAAAGCTATTTGACACTTTTAAAAAAATTAAGAAACATTAAAGSAATTAATT	900
Oy	967	GACAGATTTAATGAATTTAAGAAATCCTCGGCCAGCCAAGCTCTGGGACACCCCTATACAG	1026
Db	901	GATTAAGATTAATGAATTTAAATAATCCGCCACCGCCAAATTCGGAATTAACACCAATTAAT	960
Oy	1027	CTGCTGGACAAAGACAGAGATTAAGAGGAGCACAGAAAGAGTCAAGAGAGTGGCCAAA	1086
Db	961	CTCCTGTATTAAGACAAAAAATCGAGGACAGGAAAAAGAAATTAAGAAATTTCCAAA	1020
Oy	1087	ACCAATTAAGTCAACATAGATTTCTCTCTTACTGATCCCTTGAAGCTGGAGTACTACTTG	1146
Db	1021	ACTATTAATTAATATTTGATGATTTATTTACTTACGATCCACTGAATTTGAATTAATTAAT	1080
Oy	1147	AGAGAGAGAAATAGAAATTAAGACATCTCCGCCAAAGTGGAGACAAAGGAATCAACGGA	1206
Db	1081	AGAGAAAAAATAAATAATTAATGATTAAGATGCAAGAGGTTGAACAAAGGAATCAACTGAA	1140
Oy	1207	CCTAATGAATATCCCAATGGTGGAGGTACCCCTGCTCTTAATAGATATTCACCAAGCT	1266
Db	1141	CCCATGAATATCCCAATGAGAGTACTTATCCCTTGTCAATTAATAGATTAATTAACATCT	1200
Oy	1267	CTCAACGAGCTCAATACTCTTCGCTGACTGATTAACCCCTTGATTAATACGAAGAACCC	1326
Db	1201	TTAATGACTTAATCTTTTGGGATTTAATTAATCACTTGTGATTAATCAAAAGAACCA	1260

QY	3788	TCACCGGTGAGGCGTGCACCTCTCCGTAATGATTAACATTTCTGTCACAAATCGGAAGC	384.47
Db	481	TACACGAGAGACGATACCTCCTCCGTAATGATTAACATTTCTGTAATTTGAAATG	540
QY	3848	AATACGAGTGCCTCATCTGAAAACCTCTGGCAGCGGCTGTATAGTCTCTCAAGAAACAGC	390.7
Db	541	AATATGAGGTGTTTATATTTTAAAAACCTTTTAGCGAGCTGTTATGAAAGTTTAAAAACAAT	600
QY	3908	TGCGAGATTAACGATGATGACCTTCAATGTCAAGGTGAAGGACATCTCTGAAACGCCCTTTA	396.7
Db	601	TAGAAATTAACGTTATATGACCTTTAATGTTTAAGCTTTAAGCATATTTTAAATTCACGATTTA	660
QY	3968	ATAAGAGAGAAAAATTTCCAGAACGCTTGGAGAGCGACTTGATTCCTATATAAGACCTGA	402.77
Db	661	ATAAACCTGAAAAATTTCCAAAAATGTTTTGAAATCAGATTTATATTCATATATAAGATTTAA	720
QY	4028	CCCTCCTTACTACTAGCTGTGTCAAGACCCTATACAGTCTCTCAATATAAGAGAGAGGATA	408.7
Db	721	CATCAAGATTAATATGTTGTCAAAAGATCCATATAATTTCTTATATAAGAAAAAAGATTTA	780
QY	4088	AATTTCTGTGTAGTTTACACATATATCATAGAGACTCCATCCACACCATATATCATTTTGCCTA	414.7
Db	781	AATTTCTTAAGCAGATTAATATTAATTAATTAAGGATTCATATAGCATATATAATTTTGCA	840
QY	4148	ATGATGTGCTGGGTTATTAACAATATCTGTAGGAAAAATATCAAGTCTGACCTTGACTCTA	420.7
Db	841	ATGATGTCTTGATTAATTAATAATATTAATCCGAAAAATATATAACAGTTTGAATTTAA	900
QY	4208	TTAAAAAGTATATCAACGATATACAGGAGGAGAGATGAAAAATATGTGCTTCCGATATA	426.7
Db	901	TTAAAAATATATCAACGACACAAACAGTGAATATGAGAAATACGCTCCCTTTTAAACA	960
QY	4268	ACATCGAAACCCCTGTACAGACAGTGACAGACAAAAATGCACTCTTTCGTAATTCACCTGG	432.7
Db	961	ATATTTGAGACCTTATATAAAAAAGTTAAATGATTAATTAATTTATTTGTAATTCATTTAG	1020
QY	4328	AGGCCAAGGTCCCACTATATCTTACGAGAGAGCAATGTGGAAGTTTAAATCAAGAGAGC	438.7
Db	1021	AAGCAAAAGTTCTTAATTTATATACATATATGAGAAATCAAAAGTTAAATTAATAAGAAC	1080
QY	4388	TGAACCTACTCAAAAACAAATCCAAAGACAGCTGGCAGATTTCAAGAAAAATAACATTTTCG	444.7
Db	1081	TTAATTTACTTAAAAACAAATTCAGACAAATATGGCAATTTTAAAAAAATAACATTTTCG	1140
QY	4448	TCCGAAATGGAGACCCGTCTACGATTTATTAACACAAACATCTCCGACCAAGTTTCGT	450.7
Db	1141	TTGGAAATGGCATTTTATCAACAGATTTATACCATATATACTTATTTGACAAAGTTCTCTTA	1200
QY	4508	CCACTGCGATGTGTTTGCAAAACCTCGCCAAAACAGTGTGAGCAATCTGTCTGACGGCA	456.7
Db	1201	GTACAGGATATGTTTGGAAAATCTTGTCTAAACCGTTTATCTATTTACTTGTATGAA	1260
QY	4568	ACCTTCACAGGCGATCTGACATCTCCACAGCAATATGCTGAAGAAACAGTGTCCCCCGA	462.7
Db	1261	ACTTGCAGAGTATGTTTAAACATTTTCAACACCCAAACGCTTAAAAAACATATGTCACAAA	1320
QY	4628	ATACCGGCTGTTTACGACCTCTGAGAGAGGCGCAAGATGCAAGTGTCTCCCTGACATCTCA	468.7
Db	1321	ATTTCTGATGTTTACAGCATTTTGAATGAAAAGAGAGAAATGTAAATGTTTATTAATTTTCA	1380
QY	4688	AACAAGAGAGATTAAGTGTGTGGAGAACCCAAACCCCTACTGTGCATGTAAAACAATGTGCG	474.7
Db	1381	AACAAGAGAGATTAATATGTTTGGAAAATCCAAATCTCTACTTGTATACGAAATAATGTGTG	1440
QY	4748	GGGTGACCGCGATGCTAAATGCAACCGAGAGAGACAGCGGCTCTTAACGGAAGAATAATCA	480.7
Db	1441	GATGTGTATGACAGATGCCAAATGTACGGAAGAAAGATTCAGGTGTGCAACGGAAGAAATCA	1500
QY	4808	CATCGAGTGTACTTAACCCGAGCTCTATATACATCTCTGACGCGGATTTTTCCTCCACACT	486.7
Db	1501	CATGTGATGTACTTAACCTGATTTCTTATTCACATTTTGCATGTGTAATTTTCTGTCAAGTTCT	1560

QY	4868	CTAATTTCCTGGGACATCCGTCGCGATGATCTGATCCGTCACAGCTTCACT	4927
Db	1561	CTAAGTCTTTGAGGATATCATCTTATTAATTAATCATGTTAATATATATACAGTTTCATT	1620
QY	4928	AATAGATGATG 4939	
Db	1621	AAAAAATGTAGG 1632	
RESULT 19			
PFAMSP808			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
COMMENT			
FEATURES			
SOURCE			
gene			
CDS			
BASE COUNT			
Query Match	18 44:	Score 910 4:	DB 3: Length 1636:
Best Local Similarity	72 44:	Pred. NO. 3.3e-189:	

Matches 1181; Conservative 0; Mismatches 451; Indels 0; Gaps 0;

```

OY 3308 CCCCCTGAAGACTCTCCCGAGAGAGATCCAGACCGAGATTAACCTACGCGCAGCTCG 3367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CCCCCTGAAGACTCTTAAGTGAAGATCAATTCACAAACAAATATATATGCGCAGTTAG 60
OY 3368 AGAATCTCAAGCTCTGCTGTAAGCTGCAAGGACGACGTAAGACAAACGATCGGAGA 3427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 AAAACTTAAAGTATTAAGTAAATTAAGAAAGAAATTAAGATATATTTAAATTTAGAAA 120
OY 3428 AGAAGAGCTCAGTACCTCTCTAGCGGACGTGATCACCCTGATCGCGAGCTCAAGAG 3487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 AGAAAAAATATATCACTATATCAAGGATTCATCATATTTATTTGCTGAATTAAGAAAG 180
OY 3488 TCATTAAGACAAAGAACTACACCGGCAATAGCCCAAGCAGAAATATACAGCTGAATA 3547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TAAATAAAAATTAATAATATACAGGTAATTCACAAAGTCAAAATATATAGGATGTTAA 240
OY 3548 ACCGACTGGAATCTTACAAAGAGTTCTGCTGGAAGGAACAGATGCGCAGCTGCTGTGT 3607
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 ATGCAATTAAGATCTTACAAAAAATTTCTCCAGAAAGAACAGATGTTGCAACAGTTGTA 300
OY 3608 CTGAATCTGGCTCGACACACTGAGAGCTCAACTAAGAGCCTGCTACTACTATG 3667
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GTGAAGGTGATCGACACATTAAGAAACAAAGTCAACCAAGAAACAGCATCATCATG 360
OY 3668 TCGGAGCCGAGTCAATCAATTAACACATCTCAAGACGTGACAGTGAAGTGCATGACG 3727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 TAGGAGCAGAGTCTAACCAATTAACACATCAAAATGTCGATATGAAGTGAAGTGAAG 420
OY 3728 TCATATTTGCTGCTCTCTGCGGAGAGAGAGAGACTAGTACCTGCGGCGAGGTGG 3787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 TAATCATATAGTACTTATTTTGGAGATCCGAAAGAAATATATATGTTAGACAAGTAG 480
OY 3788 TCACCGGTGAGGCTCTCACTCTCTCGTGATTAATCACTCTGTCACAAATCGAAGACG 3847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TAACGAGGAAAGCAGTAACTCTCTGATTAATCACTCTTCTTAATTTGAAGAAATG 540
OY 3848 AATAGAGAGTGTCTATCTGAAACCTGCGAGGCGTCTATGAGTCTCCAGAAACAGC 3907
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 AATATAGAGTCTTATATTAATTAACCTTTAGCAGGTGTTATAGAACTTAATAAACAAT 600
OY 3908 TGGAGAAATACGTGATGACCTTCAATGTCACAGTGAAGACATCTGAACAGCCGCTTA 3967
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 TAGAAATTAACGTATGACATTTAAATGTTAAGTAAAGATTTTAAATTCACGATTTA 660
OY 3968 ATAAGAGAAATTTCAAGAGCTCTTGAGAGCGACTGTTCCCTATAAGACCTGA 4027
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 ATAAACGTAATAATTTCAAAAAATGTTTGAATCAAGATTTAATCCATTAAGATTTAA 720
OY 4028 CCTCCTTAAGTCTGTCAGAGCCCATACAGTCTCTCAATTAAGAGAGAGGATA 4087
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CATCAAGTATATATGTTGTCAGAGATCCATTAATTAATTTTAAGAAAAAGAGATA 780
OY 4088 AATTTCTGTAGTATCAACTATATCAAGGACCTCCAGACACCGATCAATTTTCGTA 4147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 AATCTTAAGAGTATATATTAATTAAGATTAAGATTAAGATTAATTTTGGCA 840
OY 4148 ATGATGCTGCTGGGTTATTAAGAGATCCGAGGAAAAATACAAGTCTGACTGACTTA 4207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 ATGATGCTCTGGATTAATTAATAATTAATTCGAAAAATATTAATCAAGATTTAGATTCA 900
OY 4208 TTAAGAAATATCAAGCTTAAGAGGAGAGATGAAGAAATATATGCGCTCCGAGATA 4267
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 TTAAGAAATATATCAAGCAGCAAGAGGAGAAATGAGAAATACCTTCCCTTTTAAACA 960
OY 4268 ACATGAAACCTCTGTAAGAGAGTGAACGACAAATGACCTCTCTGTAATTCACCTGG 4327
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 AATATGAGACCTTATATTAAGACAGTTAATGATAAATGATTTATTTGTAATTCATTAG 1020
OY 4328 AGCCCAAGTCTCTCAATATATCTTACGAGAGAGCAATGTGAAGTAAATCAAGAGAC 4387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 AAGCAAAAGTCTTAATATATATATGAGAAATCAACGTTAGAGTTAAATTAAGAAAC 1080

```

```

OY 4388 TGAACCTACCTCAAAACCAATCCAGAACAGCTGGCAGATTTCAAGAAAAATACAAATTTG 4447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 TTAATTAATCTTAATAAACCAATTCAGACAAATTTGGCAGATTTTAAAAAATTAACAATTCG 1140
OY 4448 TCGGAATTCAGACCTGCTCTACCGATTAATTAACCAACAACTCTCGACCAAGTTCTGT 4507
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 TTGGAATCTGATTTATTAACAAGATTAATTAACCATTAATTAATTAATGACAAAGTTCCTTA 1200
OY 4508 CCAGTGGCAGTGTGTTGCAAAAACCTCGCAAAACAGTGTGAGCAATGCTGTCAGAGGCA 4567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1201 GTACAGGTATGTTTTTGAATAATCTTGCAAAACCGTTTTATCTAATTAATCTGATGAGAA 1260
OY 4568 ACCGTCAGGCAATGCTGCAACATCTCCAGCACCAGTATGCTGAGAAAGAACGTCGCCAGAA 4627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1261 ACTTGCAAGGTATGTTAAACATTTTACAAACACCAATGCGTAAAAAAACATGTCACAAA 1320
OY 4628 ATAGCGGCTGTTCAAGGCTCTGAGAGCGGAGAGTCAAGTGTCTCTGAAGTACA 4687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1321 ATTCGTGATGTTCTAGACATTTAGATGAAGAGAAAGATGTAATGTTTATTAATTTACA 1380
OY 4688 AACAGAAAGAGATTAAGTCTGAGAGAACCCAAACCTTACCTGCAATGAAACCAATGGCG 4747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1381 AACAGAAAGGTATTAATGTTGTTGAAATCCAAATCTCTACTGTAACGAAATTAATGTTG 1440
OY 4748 GGTGTGAGCCGATGCTAAATGACACCGAGAGAGACAGCGGCTCTTAACGGAAGAAATCA 4807
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1441 GATGTGATGTCAGATGCCAATTAACCATGCAAGAAATGATCAAGTGAAGCAAGAAATCA 1500
OY 4808 CATGCGAGTGTACTTAAGCCGACCTCTATCCACTGTCGAGCGGATTTTCTCTCAGCT 4867
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1501 CATGTGAATGATCTTAACCTGATTTCTATCCACTTTTGCATGATGTAATTTCTCAGTTCT 1560
OY 4868 CTAATTTCTGTGGCATCTCTCTCTGCTGATCTCATATGCTGATACAGCTTCATCT 4927
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1561 CTAACTCTTGAAGATTAATCACTTATTAATCACTGTTAATTAATTAATTAATTAATTA 1620
OY 4928 AATAGATGAGT 4939
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1621 AAAAAATGTAGG 1632

```

RESULT 20
PFAMSP828
LOCUS
DEFINITION
P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.
ACCESSION
D13360.1 GI:391798
VERSION
EGF-like domains; major merozoite surface protein precursor.
KEYWORDS
Plasmodium falciparum (isolate 828) merozoite, DNA.
SOURCE
ORGANISM
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 1636)
Jongwutives,S., Tanabe,K. and Kanbara,H.
Amino acid sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
JOURNAL
MEDLINE
93295445
REFERENCE
2 (bases 1 to 1636)
Jongwutives,S.
Direct Submission
Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)
COMMENT
Submitted (05-OCT-1992) to DDBJ by:
Somchai Jongwutives
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi

RESULT 21
PFAMSP837 1636 bp DNA linear INV 29-MAY-2002
LOCUS P. falciparum DNA for the precursor to the major merozoite surface
DEFINITION proteins, C-terminal.
ACCESSION D13362.1 GI:391806
VERSION D13362.1
KEYWORDS Egr-like domains; major merozoite surface protein precursor.
SOURCE Plasmodium falciparum (isolate 837) merozoite, DNA.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Hemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 1636)
AUTHORS Jongwutives, S., Tanabe, K. and Kanbara, H.
TITLE Sequence conservation in the C-terminal part of the precursor to
the major merozoite surface proteins (MSP1) of Plasmodium
falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE 93295445
REFERENCE 2 (bases 1 to 1636)
AUTHORS Jongwutives, S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical
Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4
Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),
Fax:0958-47-6607)
COMMENT Submitted (05-OCT-1992) to DDBJ by:
Somchai Jongwutives
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.
FEATURES
source location/Qualifiers
1. 1636
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="merozoite"
1. 1622
/gene="MSP1"
<1. 1622
/gene="MSP1"
/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BAA02623.1"
/db_xref="GI:391807"
translation="PLKLTSESIOTENNTASLENFKYLSKEGLKLNLEKKLS
YLSGSLHLIELKEVINKNNTGSPSVNMTDVNNALLESYKFLPESTDVATVSES
GSDLEQSPKPASTHVGAESNTITSONVDEDDVITVPESEEDVDLQVY
TGEAVTPSIDNLSIKENEYELTKPLAGYRSKLENNVMTFNVNYKDIANSR
ENKRENFNVLESDIIPYKDLJSSNYVKKPKLNKKRDKFLSSYVINDSIDTL
NEANDVGYKILSEKYSKSDLSIKKYNDOGEKELPLNINIEIKYVNDKIDL
FVHLEKAVLNTTTEKSNVEYKIRKELNLTQDLAFKRNKNNVGTADISDHNH
NLTKPLSTGAVFENLAVTSLNLDNLQMLNLSOHQVKKKQCPQSGFRHDER
BECKCLNFKQEGDKCVENPPTCNENNGGDADAKCTEEDSGSNKKITCECTPDS
YPLFDIGICSSSNFLGISFLILMLILYSFI"
BASE COUNT 694 a 199 c 239 g 504 t
ORIGIN
Query Match 18.4%; Score 908.8; DB 3; Length 1636;
Best Local Similarity 72.3%; Pred. No. 7.5e-189;
Matches 1180; Conservative 0; Mismatches 452; Indels 0; Gaps 0;

QY 3308 CCCCCTGAGAGCTCTCCGAGAGAGAGATCCAGACCGAGATACCTACGCCAGCTCG 3367
DB 1 CTCACCTTAAACCTTAAGCTGAAGCAATTCACAGAAATATATATGCGCAGTTTG 60
QY 3368 AGAAGCTTCAAGGTCTGTAGCTGCAAGCAAGCTGAAGCAACCTGAACCTTGAGA 3427
DB 61 AAAACCTTAAAGTATTAAAGTAAATTAAGGAAAATTAAGATTAATTAATTTAGAAA 120
QY 3428 AGAAGAGCTCAGTACCTCTAGCGAGATGATCACCCTGATCGCGAGCTCAAGAG 3487
DB 121 AGAAAAATTAATCACTTATCAAGATGATTCATCATTTAATTCGTGATTAAGAG 180
QY 3488 TCATTAGAACAAAGAACTACACCGCATATAGCCCAAGCAGATATATACAGAGCTA 3547
DB 181 TATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240
QY 3548 ACGCACTGATCTTCAAGAAAGTTCCCTCGAAGAAAGATGTCGCGACCTGCTGT 3607
DB 241 ATGCATTAAATCTTCAAAAAATTTCTCCAGAGAAAGATGTTGCAACAGTTGTA 300
QY 3608 CTGAATCTGGCTCCGACACAGCTGAGACAGCTCAACCTTGAAGACCTGATCTATC 3667
DB 301 GTGAAGTGCATCCGACACATTAGAACAAAGTCAACCAAGAAACCAAGATCACTG 360
QY 3668 TCGAGCCGAGTCCATACATTAATACCATCTCAGAACCTCCAGCATGAGTCGATG 3727
DB 361 TAGGAGCAGAGCTTAACAAATTAACAAATCAAAATGTCGATGGAAGTATGAGAG 420
QY 3728 TCATCATTTGCTATCTTCGCGAGAGAGAGAGTACGATGACTCGCGCAGCTG 3787
DB 421 TATCATTAATCTATATTTGGAGATTCGAGAGATATATGATTTTGAAGCAAGTAG 480
QY 3788 TCACCGTGAAGGCTGACTCCTCCGATGATTAACATTCGTCCAAAATCGAGAC 3847
DB 481 TAACAGAGAGAGAGTAACTCCCTCGTAAATGATTAATCTTAAATTTGAAGAG 540
QY 3848 AATGAGAGTGCCTATCTCGAAGAACTTCGGAGGCTCATATAGTCTCAAGAAACAG 3907
DB 541 AATGAGGTTTATATTTTAAACCTTAGCAGGTTTATGAACTTTAAAGAAACAT 600
QY 3908 TGGAGATATACGTATGATCAATCTCAATGTCAGTGAAGACATTCGACAGCGCTT 3967
DB 601 TAGAAATTAAGCTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
QY 3968 ATTAGAGAGAAATTTCAAGAACGTTTGGAGAGCAGCTTGATTCCTTTAAAGAC 4027
DB 661 ATAAAGTCAAAATTTCAAAATGTTTGAATCAATTTAATTCATTTAAAGATT 720
QY 4028 CCTCCTTAAGTACGTTGTCAGAGCAGCCATACAGTATTCCTCATTAAGAGAGG 4087
DB 721 CATCAAGTATTAATTTGTCAGAGATCCATTAATTTCTTAATTAAGAAAAAGAG 780
QY 4088 AATTTCTGTAGTTACAATATATCAAGAGCTCCATCAACCGATATCAATTTGCT 4147
DB 781 AATTTTAAAGAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTTTGA 840
QY 4148 ATGATGTGCTGGGATATTAACAAGATCTGAGGAAATTAACAAGTGTGACTG 4207
DB 841 ATGATGTGCTGGGATATTAACAAGATCTGAGGAAATTAACAAGTGTGACTG 900
QY 4208 TTAAGAAATATTAACAAGTGAAGAGGAGAGAAATTAACAAGTGTGACTG 4267
DB 901 TTAAGAAATATTAACAAGTGAAGAGGAGAGAAATTAACAAGTGTGACTG 960
QY 4268 ACATCAAAACCTGTCAAGAGAGTGAAGAGAAATTAACAAGTGTGACTG 4327
DB 961 ATATTGAGAGCTTATTAACAAGTGAAGAGAAATTAACAAGTGTGACTG 1020
QY 4328 AGGCAAGTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4387
DB 1021 AAGCAAAAGTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080
QY 4388 TGAACACTCAAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4447
DB 1081 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140
QY 4448 TCGAATTCAGAGCTGTACAGATTAATTAATTAATTAATTAATTAATTAATTA 4507

Db	1141	TTGGAAATTCGATTATTAACAAGATTAATAACCAATTAATTAAGTATTTGACAAAGTTCCTTA	1200
Oy	4508	CCACTGGCATGTGTGTCGAAAAACCTCGCCAAAACAGTCTGTGCAGCATGTGCTCCAGCGCA	4567
Db	1201	TACAGAGTGATGTTTTTAAAAAATCTTGCTAAAAACCGTTTTTATCTGAATTTACTGTATGAA	1260
Oy	4568	ACCTCAGAGGATGCTGTAACATCTGCCAGAACCAATGGGTGMAAGAACAGTGGCCCCAGA	4627
Db	1261	ACTTGCAAGGATGTGTTAAATTCATTTCCACAACCATTGGCTTAATAAAAAACAATGTCCACAA	1320
Oy	4628	ATAGCGGCTGTTTTTCAGGCACTGACGAGCGCGAAGAGTGCAGTGTCTCCTGAACATACA	4687
Db	1321	ATTCGAGATGTTTTCAGACATTTAGATGGAAGAAGAAATGTAAATGTTTATTAATTTACA	1380
Oy	4688	AACCAAGAAGAGATAAGTGCCTGGAGAACCCAACCCCTACTGCAATGA AAAACAATGGCG	4747
Db	1381	AACAAGAAGGTGATTAATAGTGTGTGAAATCCAAATCTACTTGTACGAAAAATATATG	1440
Oy	4748	GGTGTAGCGCCGATCTTAATATGCAACCGAGGAAGACAGGGGCTCTACGGA AAAAATCA	4807
Db	1441	GATGTGAGTGAGATGCCAAATGTACCGAAGAAGATTCAGATGACGAAAGAAAATCA	1500
Oy	4808	CATGCGAGTGTACTATAGCCCGACCTCCTATTCACACTTTGTGAGGGATTTTTTGCTCCAGCT	4867
Db	1501	CATGTGATGTACTAATCACTGATTTCTTATCCACTTTTGATGATGTATTTTCTGCACTTCT	1560
Oy	4868	CTAATTTCTCTGGGCAATCTCTCTCTGATGATCTCATGCTGATCTGACAGCTTCATCT	4927
Db	1561	CTAATCTTCTTAGAGATATCATCTTATTAATATCATGTTAATATTAATACAGTTTCATTT	1620
Oy	4928	AATAGATCGATG 4939 	
Db	1621	AAAAATGTAGG 1632	
RESULT 22			
LOCUS	PfPMSPT9		
DEFINITION	P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.		
ACCESSION	D13356		
VERSION	D13356.1 GI:391824		
KEYWORDS	EGF-like domains; major merozoite surface protein precursor.		
SOURCE	Plasmodium falciparum (Isolate 19/94) merozoite, DNA.		
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE	1 (bases 1 to 1636)		
AUTHORS	Jongwutiwes,S., Tanabe,K. and Kanbara,H.		
TITLE	Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates		
JOURNAL	Mol. Biochem. Parasitol.	59 (1),	95-100 (1993)
REFERENCE MEDLINE	2 (bases 1 to 1636)		
AUTHORS	Jongwutiwes,S.		
JOURNAL	Direct Submission		
TITLE	Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)		
COMMENT	Submitted (05-OCT-1992) to DDBJ by: Somchai Jongwutiwes Department of Protozoology Institute of Tropical Medicine Nagasaki University 12-4 Sakamoto-machi Nagasaki 852 Japan Phone: 0958-47-2111x3747 Fax: 0958-47-6607. Location/Qualifiers 1. .1636 /organism="Plasmodium falciparum"		
FEATURES	Source		

	gene	/db_xref="taxon:5833" /chromosome="9" /dev_stage="merozoite" 1..1622 /gene="MSP1" <1..1622 /gene="MSP1" /codon_start=3 /product="major merozoite surface protein precursor" /protein_id="BA02617.1" /db_xref="GI:391825"
CDS		/translation="PLKLTSEESIOTEDNVALENFVLSKLEBKLNDNLNEKKLIS YLSGSLHLILAELEKEVINKNNTGNPSNVNTDVMNLLESKYRILPGSDVAIVSSES GSPTLEBSQRPKRPASTHYGAASNTITTSQNTDDVDVVLLIPIRGESBEDVDLGQVY TGEAVTPSIDNILSKITENEVEVLTLPLACVYSLSKQLNNWTFVNVRKDILLNR FNKNENRNVLDESLLPKRKDLTSSNYVVKDYKFLNKEERKRFLESYNYIKDSIDTDI NFANDLVGYKRLISEKYSKSDLDSIKRYINDOGKENEXELPLANNIEYLKYKPNDRID FVHLEAKVLNVTYEKSNEVAKIKEINLYLKTIOOKLADPEKNNNEVGJADISTYNHN NLILFKPSTGVFENILAKTYVISLNLDIGMLNISOHOCYKRCPCNSGCFRHLEDER EECCCLNTRKOEGRCYENPRPTCNENNNGGDALAKTEDESGSNKGKITCECTKPDS YLPFDGIFCSSNPLGISFLIILAILILYSFI"
BASE COUNT	695 a 199 c 239 g 503 t	
ORIGIN		
Query Match	18.14%	Score 908.8; DB: 3; Length 1636;
Best Local Similarity	72.13%	Pred. No. 7.5e-189;
Matches 1180; Conservative	0; Mismatches 452; Indels 0; Gaps 0;	
Oy	3308 CTCTCTGAAGACTCTCTCCGGAGAGACATCCAACCGAGATTAACGCCACCTCG	3367
Dd	1 CTCCATTTAAAACTTTAAGTAGAAGATCAATTCAACAGAGAATATTATGCGACTTAG	60
Oy	3368 AGAACTTCAGAGTCTGTCTAAGCTCGAAGCGAAGCTGAAGACAACCTGAACCTGGAGA	3427
Dd	61 AAAACTTTAAAGTATTAGTAAATTAGAAAGAAATTTAAAGATATTATTAAATTAGAAA	120
Oy	3428 AGAAGACCTGACCTCTCTACCAGACTGCATCACCCTGATCCGCCAGCTCAAGSAAG	3487
Dd	121 AGAAAAAATTTATCATCTTATTCMAAGAGATTACATCATCTTTAATTTCTGAATTTAAAGAAG	180
Oy	3488 TCATTTAAGAACAAAGACTACACCGGCATATAGCCCAAGGAGAAATATACAGACTGAATA	3547
Dd	101 TATATAAAAATTAATTAATTAACAGSTATTCTGCCAAGTGTAATTAATNOGATGTTAACA	240
Oy	3548 ACGCACCTGAATCTTTACAAGAAGTCTGCTCGCTGAAGAAACAGATGTGCGCACTGTGGTGT	3607
Dd	241 ATGCATTAGAAATCTTACAAAAAATTTCTCCCGAAGAAACAGATGTGCAACAGTTGTAA	300
Oy	3608 CTGATTTGGGCTCGGACACACTGSCAGAGTGCATCAACCTAAAGAAAGCTGCATCTACTCATG	3667
Dd	301 GTGAAGATGGATCGGACACATTAGAACAAAGTCAACCAAAGAACACAGCATCACTCATG	360
Oy	3668 TCGGAGCCGAGTCCATTAATTAACATTCACACATCTCAAGACGTGCGACGATGAGTGCATGACG	3727
Dd	361 TAGGACGACAGCTTAACCAATTAACACATCTCCAANAATGTGATGATGAAGTAGTAGACG	420
Oy	3728 TCATCATTTGTGCTATCTTCCGGAGAGCGAGAGAGACTACGATGACCTCGGCGAGTGG	3787
Dd	421 TAATCATAGTACTTATTTTGGAGAAATCCGAAGAGATTAATGATGATTTAGACACAAGTAG	480
Oy	3788 TCACCGGTGAGGCTGTCACTCTTCCGCGATTTGATTAACATTTGTGCCAAAATCGAAGACG	3847
Dd	481 TTAACAGGAGAAACCACTACTCTCTCCGTAATTTGATTAACATCTTTCTTAATTTGAANAATG	540
Oy	3848 AATAGAGTGGCTGATCTGAACCTCGGCGAGGAGGTGATATGCGTCCCAAGAAACAGC	3907
Dd	541 AATATGAGGTTTTATATTTTAAACCTTTTAGCGAGGTTTATATGAAGTTTAAAAAACAAT	600
Oy	3908 TGGAAGATTAACGTGATGACCTTCATGTGCACAGTGAAGACATTTCTGAAGACGCCCTTTA	3967
Dd	601 TAGAAATATACGTTATGACATTTTAATGTTAAGTAAAGATATTTTAAATTCACGATTTA	660
Oy	3968 ATTAAGAGAAAAATTTCAAGAACGTTCTTGAGAGCGACTGTGATCCCTATTAAGACCTGA	4027

```

Db 661 ATAAACGGAATAATTTCAAAATGTTTAGATCAGATTTATTCATATATAAGATTTAA 720
Qy 4028 CCTCTTAAGTACTGTTTCAAGAGCCATACAGTCTCTCATATAAGAGAGGATA 4087
Db 721 CATCAAGTAATTAATGTTTCTCAAGATTCATTAATTTCTTAATAAAGAAAAAGAGTA 780
Qy 4088 AATTCTCTAGTATCAACATATATCAAGAGTCCATGACACCATATTCATTTCCGTA 4147
Db 781 AATTCTTAAGCAGTTATATATATATTAAGATTCATATAGATATGAATTAATTTTGCA 840
Qy 4148 ATGATGCTCTGGGATTTATCAAGATCTGAGGAAAAATATACAGTCTGACCTTGCTTA 4207
Db 841 ATGATGTTCTTGAGATATATATAAATATATATCCGAAAAATATATAATCATAGATTAG 900
Qy 4208 TTAATAAATATATACACATTAAGCAGGCGAGATGAAAAATATATGCTCTCTGTA 4267
Db 901 TTAATAAATATATACACCAACAAAGGTAAGAAAAATATACCTCTCTTTTAAACA 960
Qy 4268 ACATCGAAGACCTGTACAAAGACAGTGAAGACAAATGACCTCTTCTGAATTCACCTG 4327
Db 961 ATATTGACACCTTATATATAAACAGTAAATGATTAATTTGATTTATTTGATTCATTAG 1020
Qy 4328 AGCCCAAGGTCCTCACTATATCTTACGAGAGAGCAATGTGGAATTAATAATCAAGAGC 4387
Db 1021 AAGCAAAAGTTCTAAATTTATACATGTGAAGAAATCAACGTAGAAAGTTAAATAAAGAAC 1080
Qy 4388 TGAATCTACTCAAAACATCCAGACAGCAAGCTGGCAGATTTTCAGAAAAATATACATTTCG 4447
Db 1081 TTAATTTACTTAAACAAATTTCAAGACAAATGGCAGATTTTAAATAAATAACAATTTTCG 1140
Qy 4448 TCGAATTTGACAGACCTGTCTACCGATTTAAACACACAAATCTCTGACCAAGTTCTGT 4507
Db 1141 TTGGAAATTTGATGATTTATCAACAGATTTATACCATTAATTAATTTGACAAAGTCTTA 1200
Qy 4508 CCACTGCGATGTTGTCGAAAACTCGCCAAACAGTCTGAGCAATGCTCGAGCGCA 4567
Db 1201 GTACAGGATGTTGTTTAAAAATCTTGTAAACCGTTTATCTAATTTACTTGTGAGAA 1260
Qy 4568 ACCTGACGGGATGCTGAACATCTCCGACAGCAAGCTGGTGAAGAAACAGTCCCCCAGA 4627
Db 1261 ACTTGCAAGGATGTTTAAACATTTTCACAAACCAATGCGTAAAAACAAATGTCACAAA 1320
Qy 4628 ATAGCGCTGTTTCAAGCATCTGACGAGCGGAGAGTGAAGTGTCTCTGCACTACA 4687
Db 1321 ATTCGATGTTTTCAGCATTTAGATGAAGAAGAAATGTAATGTTATTAATTTACA 1380
Qy 4688 AACAAAGAGAGATATAGTCGTGGAAGAACCAACCTTACTGCAATGAAGAAATGCGC 4747
Db 1381 AACAAAGAGGATTAATGTTGTTGAAATCCAAATCTTAACTGTAACGAAAAATATGCTG 1440
Qy 4748 GGTGAGCGGATGCTAAATGACACGAGAAACAGGCGCTTAACGAAAAAATAATCA 4807
Db 1441 GATGGAATGACAGATGCCAAATGACGAAAGAAATTCAGATGACGAAAGAAAAAATCA 1500
Qy 4808 CATGCGATGTAATTAAGCCGACTCTATCCACTCTTGACGAGATTTTTCCTCCACT 4867
Db 1501 CATGGAATGTAATTAAGCTGATTTCTTATCCACTTTTGAATGATTTTTCGCACTTCT 1560
Qy 4868 CTAATTTCTGGGATCTCTCTCTCTGATGATTCATGATCTGATCTGATCAAGCTTCATCT 4927
Db 1561 CTAATTTCTGGAATATATCATTTATTAATATATATATATATATATATATATATATAT 1620
Qy 4928 AATAGATGATG 4939
Db 1621 AAAAAATGTAGG 1632

```

RESULT 23
 LOCUS PFAMSP834 1636 bp DNA linear INV 29-MAY-2002
 DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

```

ACCESSION D13361
VERSION D13361.1 GI:391800
KEYWORDS EGF-like domains; major merozoite surface protein precursor.
SOURCE Plasmodium falciparum (isolate 834) merozoite, DNA.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 1636)
TITLE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE 93295445
REFERENCE 2 (bases 1 to 1636)
AUTHORS Jongwutives, S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)
COMMENT Submitted (05-OCT-1992) to DDBJ by: Somchai Jongwutives
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.
FEATURES
source
1. 1636
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="merozoite"
1. 1622
/gene="MSP1"
<1. 1622
/gene="MSP1"
/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BAA02622.1"
/db_xref="GI:391801"
/translation="PLKTSSESJOTEDNYASLENFVLKLEKTLKONLNEKKKLS
YLSGHLHIALEKVEYKKNKNTGCSNNTPDNALNEYSKKRLEPGTQVATPVSES
GSDLEOSPQRKRPASTHVGAESNTITTSQVDDDEVYITVIFGESEEDYDLAGOYV
TGEAVTSPVIDNLSIKTENEYEVLYKPLAGVYRSKLEKLENNVMTNVNKKDILNR
FNKRENFKNVLESDILPKDLSNNYVKKPKLNKLEKDKFLSSYNYTKDSIDTDL
NFANDVIGYKILSEKYSKSDLSIKYINDKQGENEYKLPNNIETLYKTVNDKIDL
FVHLEKVLNNTYKESNVEYKIKELNLYKTIDOKLADFPKNNNFVGIADLSTPDYNN
NLTKRFLSTGMVPENLAKYLYSLNLDNLOGMNIISOHCYKROCPONSGCFRILDER
BECKCLNTYKQESKCYENPNPTCENNNGGCDADAKCTEEDSGSNKKITCECTKPS
YLPEDGIFGSSSNFLGISFLILILYSL"
BASE COUNT 695 a 199 c 239 g 503 t
ORIGIN
Query Match 18.48; Score 907.2; DB 3; Length 1636;
Best Local Similarity 72.28; Pred. No. 1.7e-188;
Matches 1179; Conservative 0; Mismatches 453; Indels 0; Gaps 0;
Qy 3308 CTCCCTGAGACCTCTCCGAGGAGACATCCAGACGAGATTAACGCCAGCCTCG 3367
Db 1 CTCCATTAATAACTTTAAGTGAAGATCAATTTCAACAGAAAGATATATATGCCAGTTAG 60
Qy 3368 AGAAGTTCAAGTCTGTCTAAGCTCGAAGGCAAGCTGAAGACAACTGAACCTGAGGA 3427
Db 61 AAAAGTTAAAGTATTAAGTAAATTAAGAGAAATTAAGAGATTAATTAATTAAGAA 120
Qy 3428 AGAAGACCTGACGCTCTGACGACGACATCGATCATGCGGACGCTCAAGGAG 3487
Db 121 AGAAAAAATTTATCATACTATATCAAGTGATTAACATCATTTAATGCTGAATTAAGAGAG 180

```

OY	3488	TCCTTTAGAACACAACTACACCGGCATGCCCAAGCAGAAATATATACAGCGGATA	3547
Db	181	TAATTAATAAATAAATAATATATACAGGTAATCTCTCAAGTGAATAATATACGATTTAA	240
OY	3548	ACGCACCTGGAAATCTTTACAGAAAGTTCCTGGCTCAAGGAACAGATGTGGCACTGTGGT	3607
Db	241	ATGCATTTAGAAATCTTACAAAAAATTTCTCCAGAAAGAAACAGATGTGGCAAGTTGTA	300
OY	3608	CTGAATCTGGCTCCGACACACTGGAGCAGTCTAACCTAAGAAGCTGCATCTACTATG	3667
Db	301	GTGAAGTGGATTCGACACACTTGAACAAAGTCAACCCAAAGAAACCCAGCATCACTCATG	360
OY	3668	TCGGAGCCGAAATCCAAATACATTTACCAATCTCAGAAAGCTCGACGATGGAGTGCATGACG	3727
Db	361	TAGAGAGCAGATCTTAACACAAATAAACAACATCAACAAATGTCCATGATGAAGTATGACG	420
OY	3728	TCATCATTTGTCCCTATCTTTGGGAGAGCAGAGAGACTACGTACCTCGGCCAGGTGG	3787
Db	421	TAATCTATGTACCTATATTTGGAGAAATCCGAAAMAAATATGATGATTTAGACCAAGTACG	480
OY	3788	TCACCGGTGAGCGGTGCACATCTCCGCTGATGATTAACATCTGTGCCAAATTCGAAAGC	3847
Db	481	TACACGGAGAACGATCACTCTCTCCGTAAATTAACACTTTCTTAAATTTGAAATG	540
OY	3848	AATACGAAGTCTCTATCTGAACCTCTGGAGGCGTCTATAGTCTCTCAAGAAACAGC	3907
Db	541	AATATGAGGTTTATTTATTTAAACCTTTACAGAGTGTATATGAAGTTTAAAAAACAAT	600
OY	3908	TGGAGAAATACGTATACCTCTCAACCTCAAGCTGAAGGACATCTGAAACACCGCTTTA	3967
Db	601	TAGAAATTAACCTTATATGACATTTTATATGTTTATAGGATATTTTAAATTCACATTTA	660
OY	3968	ATTAAGAGAAATTTTCAAGACGCTTGGAGAGCGACTGTGATTCCTTTAAAGACCTGA	4027
Db	661	ATTAACCGTGAATTTTCAAAAATGTTTAAATCAACATTTAATTCATTTAAAGATTTAA	720
OY	4028	CCCTCTCACTAGTGTGAAGGACCCATACAGTCTCTCAATTAAGAAAGAGAGATA	4087
Db	721	CATCAAGTATTAATGTTGTCAAGATCCATATTAATTTCTTATTAAGAAAAAAGATTA	780
OY	4088	AATTTCTGTCTAGTTACACATATATCAAGAGCTCCATGCACACCCGATATCAATTTTCGTA	4147
Db	781	AATCTTAAAGAGTTAATAATTAATTAATTAAGATTCATTAAGATTCAGGATTAATAATTTTGC	840
OY	4148	ATGATGTCTGGGCTATTTACAAGATCCTGAGCGAAAAATATCAAGTCTGACTGCTA	4207
Db	841	ATGATGTCTGGATTTATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	900
OY	4208	TTAAAAAGTATATCAACGATTAAGCAAGGCGAGATGAATAATATGCGCCCTCGATATA	4267
Db	901	TTAAAAAATATATCAACGACAAACAAGGTGAATAATGAGAAATATACGTTTGAATTAACA	960
OY	4268	ACATCGAAACCCCTGTACAGACAGTGAACGACAAATTCGACTCTTTCGTAATTCACCTGG	4327
Db	961	ATATTTGAAACCTTTATTAATAACAGTTAATATTAATTAATTAATTAATTAATTAATTAATTA	1020
OY	4328	AGGCCAAGCTCTCAACATATATCTAGCAAGAGCAGATGGGAAGTTAAATCAAGAGGAC	4387
Db	1021	AAGCAAAAGTTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1080
OY	4388	TGAACCTACCTCAAAAACAATCCAAAGACAAAGCTGGCAGATTTCAAGAAAAATAACAATTTG	4447
Db	1081	TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1140
OY	4448	TCGGAATTCGACAGACTGTCTACGATTTATTAACCAACAATCTCCTGCACCAAGTTTCTGT	4507
Db	1141	TTTGAATTCGATTTATTAACAAGATTTATTAACCATTAATTAATTAATTAATTAATTAATTAAT	1200
OY	4508	CCACCTGGCATGTGTTCGAACAACCTCGCCAAAACAAGTGTGGAGACATCTGCTCGACGGCA	4567
Db	1201	GTACAGGATGTGTTTGAAGAAATCTTGCTTAACACGTTTATCTATTAATTAATTAATTAATTA	1260
OY	4568	ACCTTCAGAGGATCTGTACACATCTCCAGCACCAATGTGCTGAAGAAACAGTCCCCACGA	4627

DB	1261	ACTTGCAAGTATGTTAAACATTTTACACAGCAATGGCTAAAAAAGATGTCCACAA	1320
OY	4628	ATAGCGGCTGTTTCAGGCACTGACGAGCGGGAAGAGTCAAGTGTCTCGAAGTACA	4687
Db	1321	ATTCTGGATGTTTCAGACATTTAGATGTAAGGAAGAAATGTAAAGTTTATTAATTTACA	1380
OY	4688	AACCAAGAGGACATTAAGTGGCGTGGGAACCCCAACCCCTACCTGCAATGAAACATGGCG	4747
Db	1381	AACCAAGAGGATGTAAGTGGTGGTAATCTCAATCTCTGTTGTAACGAAATTAATGGTG	1440
OY	4748	GGTGTGACGCCGATCTAAATGCAACCGAGGAAGACAGCGGCTTAACGGAAGAAATCA	4807
Db	1441	GATGTGATGCAATGCAAAATGTACCGAAGAAATTCAGGTAGCAACGGAAGAAATTCAC	1500
OY	4808	CATGCGAGTGTACTAAGCCCGACCTCTATCCACTCTTGACGCGGATTTTGTCTCAGCT	4867
Db	1501	CATGGAATGATACATAACCTGATTTCTTATCCACTTTTGATGGATTTTCTGACGTTCT	1560
OY	4868	CTAATTTCCCTGGGCACTTCCTTCCTGCTGATCCCTGATGCTGTAACAGCTTCACT	4927
Db	1561	CTAAGTCTTGTAGCAATATCATCTTATTAATCTATGTTATATTAACAGTTTCATTT	1620
OY	4928	AATGATCCGATG 4939	
Db	1621	AAAAAATGTAGG 1632	
RESULT 24			
PFAMSP844			
LOCUS			
DEFINITION	P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.	1636 bp	DNA linear INV 29-MAY-2002
ACCESSION	D13363		
VERSION	D13363.1	GI:391816	
KEYWORDS	EGF-like domains; major merozoite surface protein precursor.		
SOURCE	Plasmodium falciparum (isolate 844) merozoite, DNA.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 1636)		
TITLE	Jongwutiwes,S., Yanabe,K. and Kanbara,H.		
JOURNAL	Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates		
MEDLINE	Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)		
REFERENCE	93295445		
AUTHORS	2 (bases 1 to 1636)		
TITLE	Jongwutiwes,S.		
JOURNAL	Direct Submission		
COMMENT	Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)		
COMMENT	Submitted (05-OCT-1992) to DDBJ by:		
COMMENT	Somchai Jongwutiwes		
COMMENT	Department of Protozoology		
COMMENT	Institute of Tropical Medicine		
COMMENT	Nagasaki University		
COMMENT	12-4 Sakamoto-machi		
COMMENT	Nagasaki 852		
COMMENT	Japan		
COMMENT	Phone: 0958-47-2111x3747		
COMMENT	Fax: 0958-47-6607.		
FEATURES			
Source	Location/Qualifiers		
	1..1636		
	/organism="Plasmodium falciparum"		
	/db_xref="taxon:5833"		
	/chromosome="9"		
	/dev_stage="merozoite"		
	1..1622		
	/gene="MSP1"		
	<1..1622		
	/gene="MSP1"		
CDS			

AUTHORS Jongwutives,S., Tanabe,K. and Kanbara,H.
 TITLE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
 JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
 MEDLINE 93295445
 REFERENCE 2 (bases 1 to 1636)
 AUTHORS Jongwutives,S.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4 Sakamoto-machi, Nagasaki 852, Japan (tel.0958-47-2111(ex.3747), Fax:0958-47-6607)
 COMMENT Submitted (05-OCT-1992) to DDBJ by:
 Somchai Jongwutives
 Department of Protozoology
 Institute of Tropical Medicine
 Nagasaki University
 12-4 Sakamoto-machi
 Nagasaki 852
 Japan
 Phone: 0958-47-2111x3747
 Fax: 0958-47-6607
 FEATURES
 source
 Location/Qualifiers
 1..1636
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /chromosome="9"
 /dev_stage="merozoite"
 1..1622
 /gene="MSP1"
 /gene="MSP1"
 /codon_start=3
 /product="major merozoite surface protein precursor"
 /protein_id="BAA02620.1"
 /db_xref="GI:391797"
 /translation="PKLTLSSESIOTEDNIVASLENFKVLSKLEKIKDNLEKKLS
 YLSGLHLIAELKEVINKKNTGNSPVNNDVNALESYKFLPEGDVATVSES
 GSDTLEOSQPKRPASTHVGAEENITTTSONVDEVDVIVIPFGSEEDYDLQGV
 TGEAVTPSYIDNLSKIENEYELVAKPLAGYRSIKOLENNWTFNNVNDILNSR
 FNKRENFKNVLESDLIPIKDLSSNNVVDYKPLNKEKDFSLSYNIKDSIDTI
 FNANDVLYKILSEKSKSDISIKKIINDKGENKIYPLANNIETLYKYNDKIDL
 FYIHLKAVLNTYTESNVEVIREKELNTLKTIDKLDKPKNNNEVGLADSTDYHN
 NLITKFLSTGVFENILAKTVLSNLDGLNGLMLNLSQHCVKKQCPQNSGCFRHLDER
 BECKCLANKQEGSKCVENPNTCENNGGCDADAKCTEEDSGSNKKITCECTPDS
 YPLFDGIFSSSNFLGISPLILMLILVLSFI"
 BASE COUNT 694 a 199 c 239 g 504 t
 ORIGIN
 Query Match 18.3%; Score 905.6; DB 3; Length 1636;
 Best Local Similarity 72.2%; Pred. No. 3.8e-188;
 Matches 1178; Conservative 0; Mismatches 454; Indels 0; Gaps 0;

QY 3608 CTGAATCTGGCTCCGACACCTGGACAGCTCAACCTAAGACCTGCATCTACTCANG 3667
 ||||| ||| ||||| ||||| ||| ||| ||||| ||||| ||| ||||| ||||| |||||
 Db 301 GTGAAGTGTATCTCGACACATTTAGAACAAAGTCAACCAAGAACAGCATCACTCANG 360
 QY 3668 TCGGAGCCGATCCCAATPACATTTACACATCTCAGAACCTCGACGATGAGTGCATGAG 3727
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 361 TAGGACGAGGCTTAACACATTAACACATCAACAAATGCGATGATGAGTGCATGAG 420
 QY 3728 TCATCATTTGGCTATCTCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3787
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 421 TAATCATAGTACTATATTTGGAGATCCGAGAGATTTATGATGATTTAGGCAAGTAG 480
 QY 3788 TCACCGGTGAGGCTGTACACTCCCTCCGTATGATTAACATTTCTGCCAAATGAGAGAG 3847
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 481 TAACAGAGAGAGAGATACCTCCGTAATGATTAACATCTCTTGAATTTGAATTTGA 540
 QY 3848 AATACGAGTGTCTATCTGAAACCTCTGGCAGGCGCTATAGGCTCTCAGAAACAGC 3907
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 541 AATATGAGGTTTATTTTAAACCTTTAGCAGGCTTTATAGAACTTTAAAAAACAAAT 600
 QY 3908 TGGAGATAACGATGACCTTCATGTCACGTTGAAGAGCATTTGACACGCCGCTTTA 3967
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 601 TAGAAATATACGTTATGACATTTAATGTTATGTTAAGGATTTTAAATTCACGATTTA 680
 QY 3968 ATAGAGAGAAATTTCAAGAACGCTTGGAGAGCAGCTTGATTCCTATTAAGACCTGA 4027
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 661 ATAAAGTGAAATTTCAAAAATGTTTGAATCAATGATTTAATTCATATTAAGATTTAA 720
 QY 4028 CCTCTCTACTAGTCTGTGCAAGAGCACCACATACAGTTCCTCAATAAAGAGAGGATA 4087
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 721 CATCAAGTATTTATTTGTCAAAGATCCATATTAATTTCTTAATAAAGAGAGAGATA 780
 QY 4088 AATTTCTGCTAGTCAACATATATCAAGAGCTCATCGACACCGATATCAATTTGCTTA 4147
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 781 AATTTCTAGACATTTATATATATATATAGATTCATATAGATTAATAATTTTGCA 840
 QY 4148 ATGATGCTGGGGTATTTACAGATCTCGAGCAAAATTCAGAGTTGACCTTGACTCTA 4207
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 841 ATGATGCTGGGGTATTTATTAATTTATTTCCCAAAATTTAATTCAGATTTGATTTCA 900
 QY 4208 TTTAAAGTATATCAAGATTAACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4267
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 901 TTTAAAGTATATCAAGATTAACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 QY 4268 ACATCGAAACCTGTACAGAGACAGTACAGACAAATTCAGACCTCTCTGTAATTCACCTG 4327
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 961 ATATGAGACCTTATTAACAGAGTAAATGATTAATGATTTATTTGTAATTCATTTAG 1020
 QY 4328 AGGCCAAGGTCTCAACTACTTACGAGAGAGCAATGTGGAAGTTAAATCAAGAGAGC 4387
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1021 AAGCAAAAGTCTTAATTTATCATATGAGAAATCAAAACGTAGAGTTAAATTAAGAGAGC 1080
 QY 4388 TGAATCTCAAAACAAATCCAGAGACAGTGGCAATTTCAAGAAATTAACAAATTTG 4447
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1081 TTAATTTCTTAATAAACAAATTCAGAGCAAAATGTGCAATTTTAAAAAATTAACAAATTTG 1140
 QY 4448 TCGGAATGAGAGCTGTACCGATTTATACCAACAACATCTCGACCAATGCTTGT 4507
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1141 TTGGAAATTTGCTGATTTATACAGAGATTTATACCATTAATTAATTTATGACAAAGTTCTCTTA 1200
 QY 4508 CCACGTGATGTGTTTCGAAAACTCGCAAAACAGTGTGAGCAATCTGTGAGAGGAG 4567
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1201 GTACAGATGATGTTTGAATAAATCTGTCAAAACGTTTATGATTAATTTACTGATGAGAA 1260
 QY 4568 ACCTGAGGAGATGCTGAACATCTCCAGACCAATGCGTGAAGAAACAGTCCGCCAGAG 4627
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1261 ACTTCAAGAGTATTTAAACATTTCAACACCAACCAAGCTTAAAAAACAATGTCCCAAA 1320
 QY 4628 ATAGCGGCTGTTTCAGGCACTCTGAGAGAGCGCAAGAGGCAAGTCTCCGAATTTACA 4687
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1321 ATTTCTGATGTTTCAGACATTTATGATGAAGAGAAAGATGTAATGTTTATTAATTTACA 1380

[illegible][illegible]

Db 1036 TTGAGAAAGACACATATATTTAATAGCGTTTGTAGAAAAACGTAATTGACACTTTAAAAA 1095
 QY 943 AATGAAATATTCAAAGAACTGCTGCACAGATTAATGAATTAAGA---ATCCGCGCCA 999
 Db 1096 AATGAAATATTCAAAGAACTGCTGCACAGATTAATGAATTAAGA---ATCCGCGCCA 1155
 QY 1000 GCCAACTCTGGGAACACCCCTAAACGCTGTGACAGAAAGAAATGATGAGAGAC 1059
 Db 1156 CCCACTACTGAAAGTAAACCAATACTCTCTGATNAGAACAAAAATCGAGGAACAC 1215
 QY 1060 GAGAAAGAGTCAAAAGATCGGCAAAACGATTAAGTCAACATAGTCTCTCTTACT 1119
 Db 1216 GAGGAAAAATTAAGAAATGCGAAAACTTAATTAATTAACATGTAAGTATTAATTA 1275
 QY 1120 GATCCCTTGGAGTGAAGTCTACTTGAAGAGAGATTAAGATTAAGATTCCTCCGC 1179
 Db 1276 GATCCACTTGAATTAAGATTAATTAAGAGAAAAATTAAGATTAAGATTCCTCCGC 1335
 QY 1180 AAGATGCA-----GACAAAGAAATCAACCGAACCTTAATGATTCCTCAATGCTGTG 1230
 Db 1336 AATTCACAGAGATCTCAAGAAATCTGTTCAAAATACCAAAAGTCTTAATCCAAATGTAT 1395
 QY 1231 AGCTACCCCTGCTTATTAAGATTAACCA-----GGCTCAACGAGCTCAAT 1281
 Db 1396 GTATATCTTACCCTACTGCTGATTAATTAATTAATTAATTAATTAATTAATTAAT 1455
 QY 1282 AGCTGCGTGAATTAACCCCTTGAATTAACGAAAGAACCTTAAAGATTAATTAAT 1341
 Db 1456 TCATATGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1515
 QY 1342 ACAGACATGAGAGAAAGATTAATTAACGAAATTAAGAGAGATTAATTAATTAAT 1401
 Db 1516 GATTAATTAAGAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1575
 QY 1402 AAGAAATTAAGAGATTAAGAGAAAGATTAAGAGAGAGATTAATTAATTAATTAAT 1461
 Db 1576 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1535
 QY 1462 ACTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAAT 1521
 Db 1636 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1695
 QY 1522 GACCTGACCACTTCCGAAAGATTAAGAGAGAGATTAAGAGATTAATTAATTAAT 1581
 Db 1696 GACCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1755
 QY 1582 ACACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1641
 Db 1756 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1815
 QY 1642 AAGCTCTTAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAAT 1701
 Db 1816 AAGCTCTTAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAAT 1875
 QY 1702 AAGTATTAAGATTAAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAAT 1761
 Db 1876 AATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1935
 QY 1762 ATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 1821
 Db 1936 ATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 1995
 QY 1822 GATGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 1881
 Db 1996 AATG---CTTCTTAAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2052
 QY 1882 ATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 1941
 Db 2053 ATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2112
 QY 1942 CATTAATTAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2001
 Db 2113 GATGATTAATTAAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAATTAAT 2172

QY 2002 ATGTACTCAAGAAAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2061
 Db 2173 ATGTAATTAAGAAAGAGATTAATTAAGAGATTAATTAATTAATTAATTAATTAATTAAT 2232
 QY 2062 ATCAACGAGAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2121
 Db 2233 TTAAGAA--AGAAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2290
 QY 2122 ACAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2181
 Db 2291 AATCAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2249
 QY 2182 CTCGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2241
 Db 2350 GTAAAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAATTAATTAAT 2409
 QY 2242 CCAAGTCCCTTCCAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2301
 Db 2410 CCATCAACACCAAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2469
 QY 2302 AAGACGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2361
 Db 2470 AATTCACAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2529
 QY 2362 TCCTACATTCGCAAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAATTAAT 2421
 Db 2530 TCATTAATTAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2589
 QY 2422 AATCAATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAAT 2481
 Db 2590 GAGTATTAATTAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2649
 QY 2482 CTCGCTTCAATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAATTAAT 2541
 Db 2650 TTAATTAATTAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2709
 QY 2542 AATTCCTCTCAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAAT 2601
 Db 2710 AATGATTAATTAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAAT 2769
 QY 2602 AATCTCAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2642
 Db 2770 AATCTCAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAATTAATTAATTAAT 2810

RESULT 27
 PPAP190A 3518 bp mRNA linear INV 26-APR-1993
 LOCUS P.falciparum merozoite 190 kd precursor protein (p190) mRNA, 5'
 DEFINITION
 accession M35727.1 GI:160549
 version 1
 keywords p190 gene; variable surface antigen.
 source Plasmodium falciparum (individual isolate RO-33 Ghana) cDNA to mRNA.
 ORGANISM
 Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE
 1 (bases 1 to 3518)
 Certa, U., Rotmann, D., Matile, H. and Reber-Liske, R.
 A naturally occurring gene encoding the major surface antigen precursor p190 of Plasmodium falciparum lacks tripeptide repeats
 EMBO J. 6 (13), 4137-4142 (1987)
 JOURNAL MEDLINE
 PUBMED 8816657
 FEATURES
 source 3327688
 location/Qualifiers
 1..3518
 /organism="Plasmodium falciparum"
 /isolate="RO-33 Ghana"
 /db_xref="taxon:5833"
 1..>3518
 /product="p190 protein"
 /note="RO-33 mRNA"

mrna	55. >3518	/product="p190 protein"	/note="Camp mrna"
mrna	203. >3518	/product="p190 protein"	/note="MAD20 mrna"
mrna	266. >3518	/product="p190 protein"	/note="K1 mrna"
CDS	337. >3518	/codon_start=1	/product="p190 protein"
		/protein_id="AA29715.1"	/db_xref="GI:160550"
		/translation="MKIIFPLCSFLPIITMQCTHSTHGYELKKEALEDAVITGS LFOEKWYLDKQANTQYVAPRADAVSTQSKANPNPAGTSTKAIKSPGAPDS DDSDSDAKSTADLKHRVQVLFETIKELKPELEDTLNMHLLTCDNIHGKYLIDGE EINILKLYNEFDPLKLANDVANDYCOIIPNLKIRANELDYLKLVGYRPLDF IKDWGMEDEYIKKNTTIANINIELIGSKRTIDQNKANDEEKKLYOAOYDLFY NKOEAHNLISVLEKRIIDLTKKENKILKLEDDIKIDAERTTGTVNOILSLREK ESROEEKIKELAKTIKRNIDRLFPDLELYLREKKYADMPKQSPKSVQIPV PYNGIYPLPLPIDIHNSLADNDKNSYGLMDPHREKINEKITDINKERKIFINNI KKQJIDLEKNNHNTKEONKKLLEDYERKSKDELELKEYLEMENNENKNDVVDKTS ARYYNVEKORYNNKFSSSNNSYVYVOKLKAUSYLEDYSLRGIISERDNHYTLT GLEADIKRLTEIEKSENKILIEKNFKGLTHSANSLEVDIVKYOQVLLIKIEDL RKILDFLKNALKDSIHVNIHPQNPPEYVLYLVLEKVDLKEIFPKQMDMLKEQ AVLSITQPLAAESETTEDGSHHTLSQSEGETVEETEPEYGHHTVYLLPPEKV KVVNSIIEHKSNDNSQALTITVYIKLIDELFTSYICHHYIVNSMDQKILEVYL TPRENELEKSCDRDLFNIONINIPAMSLYDSNMNDLQHLFPELYKEMLYTLKLE ENHITKLEEDKPLTIGTSSTSSPGNTLVNTAQSATHSNSONQGSMASTYQNGVA SGPRAVESHDDPLTVLISINDLGLVSLNIGNTKVPNDLITSTEMEYENILKI MIFPNDIDIKQFVNSKSRVITGLTEQKNALNDEIKRLKIDPLQISFDLYKRYKILADR LFNKKELDGQDMQIKKTLTKLEQLESKINS"	
misc_feature	1306. 1347	/note="this frameshift results in 11 new amino acids not present in MAD-20 or K1; frameshift mutation PS-1"	
misc_difference	2496	/note="D1 in MAD-20 mrna"	
misc_difference	2544	/note="D2 in MAD-20 mrna"	
misc_feature	3247. 3270	/note="results in a single point mutation shifting reading frame to K1; frameshift mutation PS-2"	
BASE COUNT	1570 a 430 c 437 g 1081 t		
ORIGIN			
Query Match	14.4%;	Score 711.8;	DB 3; Length 3518;
Best Local Similarity	56.8%;	Pred. No. 1.6e-145;	
Matches 1511; Conservative	0;	Mismatches 1072;	Indels 76; Gaps 8;
OY	10	ATGAAATCATTTTCTTCCTCTGTTCAATTTCTGTTTTATCATCATATCTCAGTGC	69
DB	337	ATGAAGATCATATCTTTTATATGTTCAATTTCTTTTATATATAAATCAACAATGTG	396
OY	70	ACCCAGCAATCCATCAGCAGCGTGGTTAAGAAACGGAAGCGTTGGAGATGACGCT	129
DB	397	ACACATGAAAGTATTCACAGAAGCTGTCAAAAACTAGAACGCTTATGAGAATGCA	456
OY	130	ACCGGATACAGCGCTTCACAGAAGGAGAAGATGCTGTAATGGAAGGACGAGTGG	189
DB	457	ACAGGTTATAGTTTATTTCAAAAGGAAAAAATGATTAAGATGATGAGCAAAAT	516
OY	190	GCCGTTACACACGACACACCGGTTCTAAAGGTCGTGCGCTAGCAGGTCGCGT	249
DB	517	GTTGTTGCAAAAGCTGCAGATGCTGTAACTCAAAAGCTTAAAAATCTCCAGGT	576
OY	250	TCTGTGAGCCTCTGGGGGTTCGTCGCTCCGCGGACGCTGAGCATCAGTGCATG	309
DB	577	ACAGTACTTCAAGTACGACGAAGTACTAAAGTGCTATTAAGATCTCCACAGTG	633
OY	310	GCAAGCGCGGTTCGCGGGAACAGTGAAGAACAATCATCTGACACTCTAGCGAT	369
DB	634	-----ATCCTTCAGATGATTTCAAGTGATTTCA 660	

OY	370	GAGCCAAAGCTCTACGCCGACCTCAAGCACCGAGTAGAAGTAATCTCTCACTATCAG	423
- Db	661	GATGCTAAATCTTACGCTGATTTAAACATAGAGTTCCAAATTTACTTGTCCTCAATTTAA	720
OY	430	GAGCGAAGTCCCGACAGGTTGTCGACCTCACAATCATATCTCGACCTGAGTGATATAC	489
Db	721	GAATCAAAATATCCCGAAGCTTTGATTTTACCACATCATATCTTAACTTTGTGTATAT	780
OY	490	ATTCATGGCTTCAAAATATCTGATTTGACGGTTACGAGAGATCAATGAACTCCTGTACAG	549
Db	781	ATTCATGGTTCAAAATATTTAATGTATGATATGAAGAAATTAATGAATTAATATATATAA	840
OY	550	TTGGAATTTCTACTTCGACTGCTTAAGGGCCAACTAATGACGTTTGGCCATATCAT	609
Db	841	TTAAACCTTTATTTTGTATTTATTAAGAGCAAAATTTAAATGAATGTATGTCTATATATAT	900
OY	610	TGTCAAATTCATTCATTTTGAAGATACAGCCACAGCTGGAGATTTATGAAGATGTG	669
Db	901	TGTCAAATACCTTTCAATCTTTAAATTTCTGCAATATGATTTAGACGTACTTAAAAACTT	960
OY	670	GTCCTTGGATATTCGCAAGCCTCTCGAACATCAGAGCAATGTGGGAAAGATGAGAT	729
Db	961	GGTTCGGATATGAGAAACCACTTAGACTTTATTAAGATATATGTAGAAAAATGGAAT	1020
OY	730	TATATTTAAAAAGATTAGAGACCCTTCGAGAACATTAAAGACTGATCGAAGATATCCAA	789
Db	1021	TATATTTAAAAATTTAAACACCATATGCAAAATATTAAGAAATTAATGAAGAAAGTATAG	1080
OY	790	AAGACCATATGACAAATATTAAGATTCGACACCAAGAGAGAAAGAAAGATGTATACAG	849
Db	1081	AAACATATGATCAAAATATGAATTCGAGATATATGAAGAAAGAAAAAATTAATATCCAA	1140
OY	850	GCCCAATACGACTGTCCATCTATTAACAAACAGCTTGAAGAAAGCCATATCCTATCAGC	909
Db	1141	GCTCAATATGATCTTTTATTTATCAATTAACAATTCAGAAAGACACATATTTATTAATGAC	1200
OY	910	GTCATGGAGAAAGCCATATGACACCCATCAAGAAATGAAGAAATATCAAGAACTGCTGAC	969
Db	1201	GTTTATGAAAAAGCTATTTGACACTTTTAAAAAATAAGAAACATTAAGAAATTTACTTAA	1260
OY	970	AAGATTATGAAATTTAAGAAATCCT--CCGCGACCCAACTCTGGGAACCCCTATACAGC	1027
Db	1261	GATATAGTAAATTTAAATTAATAGATGCCGAAAAACCATATGAGAGTGAAACCAATCTCT	1320
OY	1028	TGCTGGACAGACAGACAGAAATAGAGAGACGAGAGAAAGATCAAAAGATTCGCCAAA	1087
Db	1321	TCCCTGAGATTAAGAAAAAGGTGAG--ACGCGAAMAAAAATTAAGAAATTTGCCAAA	1378
OY	1088	CCATTAAGTTCACATGATTTCTCTCTCTATCTATGATCCCTTAGCTGGAGTACTACTTGA	1147
Db	1379	CTATTAATATTTAATCATTTGATATAGTTATTTATCTATCTACTCTTAATATAGAAATTTATTA	1438
OY	1148	GAGAGAGAAATTAAGAAATATAGACATCTCCGCCAACTGCA-----GAAACAGAAAT	1198
Db	1439	GAGAAAAAATTAAGAAAGTTGATGTATTAACACTTAATCACAGATCCTACGAAATCTGTTC	1498
OY	1199	CAACCGAAGCTAATGAATATCCCAATGTGTGACGTACCTCTGTCTTATACATATCA	1258
Db	1499	AAATATCCAAAGTTCTCTATATCAAAATGGATGTATATCTTTACACCTCACCTGATATTC	1558
OY	1259	ACAA-----CGCTCTCAGGAGCTCATAGCTTGGGTACTGTGATTAACCCCTTGC	1309
Db	1559	ATTAATTCATTAAGTCGCGAGATATGATTAATAATTCATATGTGTATTTAATGAATCTCTCA	1618
OY	1310	ATTATACGAAAGAACCCCTTAAGAAATATCTACACAGACAAATGAGAGAAAGATTTATCA	1369
Db	1619	CTAAGAAAAAATTAATGAAAAAATTAATTAACAGATTAATTAAGAAAGAAAAATATCATTTA	1678
OY	1370	ACGAATATCAAGGAGAGATCAAAATTTGAGAGAGAGAAATTTAGAGTGCAGAGAAAGTT	1429
Db	1679	ATAACATTAATAAAACAAATTTGATTTAGAGAAAAAATAATTAATCACACAAAAAGACAA	1738

OY	1430	ACGAGAGCCGACGAGAAAGCTTAACAGATATCCTAAAGAGTATGAAAGGCTGCGAAGC	1489
Db	1739	ATAAAAATTACTGGAAGATTATGAAAAAGTCAAAAAAGATTTATGAGACATTACTTGAA	17989
OY	1490	AGATCTATGATTCCTCAAAATTCACAAATPACATCGACCTGACCAACTCTCGAANAATGATGG	1549
Db	1799	AATTTTATGAAATGAATTTATATATATATTTTAAACAAGATGTCGTAGATTAATATTTCA	1858
OY	1550	GAAACGGTACTCTTACAAAGTGAGACAACCTGACACACATATATACCTTGTGACTCTATG	1609
Db	1859	GTCGACAGTATATCACTATTAATGATGAAAAACCAAGATATATATATATTTTCATCTCTCA	1918
OY	1610	AGAAATTCAAACCATATCTTGAGAGCTGCCAACACTCTTAAGTATATGAGGAGCTATT	1669
Db	1919	ATAATTCGTATATATAAGTTCCAAAATTTAAAAAGGCTCTTTCATATCTTGAAGATATT	1978
OY	1670	CTCTCGGACATCTGTTGTGAGAAAGACATAAGTATTACAAAGATCTCATACATATAGA	1729
Db	1979	CTTTAAGAAAAGGAATTTTCGAAAAAGATTTTATATCATTTATTTATCTTGAAAACCTGGCC	2038
OY	1730	TCGAAAACGAGATCGAGACCTTGTGTGAGAACATTAAAGAGATGAGAACAGTGTGTTG	1789
Db	2039	TCGAGAGCTATTTAAAAAATTTAAACAGAGAAATTAAGAGTATGAGTGAACAAATTTCTAG	2098
OY	1790	AGAAAGATATACAAAAGACGAAATTAACCCATATGAGAAATCCTGGAGGTCCTCCGATA	1849
Db	2099	AAAAAATTTTAAAGGACTACACATTCAGCAATCTCTTC ---CTTGAAGATATCTGATA	2155
OY	1850	TTGTAAAGTCCAAAGTCACAGAGGTGCTCCTCATGAAACAAGATTGATGACTCAAGAGA	1909
Db	2156	TTGTAAATTTACAGTACGAAAAAGTTTATTTATTTAAAAAATATAGAAAGCTTAAACAAAGA	2215
OY	1910	CTCAACTCATTTCTGAGAGACGTGAGTTAAACATATATATACATGTGCCGAATAGTTATA	1969
Db	2216	TAGATATTATTTTTAAAAAATGCAACACTAAACATAGTATTCATGTACCAAAATATTATTATA	2275
OY	1970	AGCAGAGGAATTAACGACAGAACCATATCTACTCTCATCTCAAGAAAGATAGACAAAC	2029
Db	2276	AACCACAAAATTAACGAAACCAATATTTATTTATTTATTAATAAAAAAGATAGATTAAT	2335
OY	2030	TGAAAGTGTTCATGCCCAAGTCGAGAGGCTGATCAACACAGAAAGAAAGACATTATAA	2089
Db	2336	TAAAGAATTTATATACCAAAAGTAAAGACATGTTAAAGAAAGACAGCTGCTATTACAA	2395
OY	2090	CTGAAGGACACTGATTAATCTCGAGCCTTCCACAGAGAAGAGATTAACCGGACAGGCTA	2149
Db	2396	GTTATTACACAACCTTTTAGTTG -----CAGCAAGCAAAACAACACTGAAAGTGGGG	2443
OY	2150	CCACCAAGCCCGGACACAGGCGGGTTCAGCTCTCGAAGGCGATATGCGTCAAGCTCAG	2209
Db	2444	GTCACTCCACACACACTTATTCACAACTGAGAGAAACAAAGAAAGAACAGAAACAGAAAG	2503
OY	2210	CACAAGAGCAGAAACGACGACAGCTTCAGTGCAGCTGCCGCTTCAGAGGCTTAAGCTC	2269
Db	2504	AAACAGTAGAGCACACAAACAACGGTAAACATTAATATACCACCAAAAAGATTAAGTGG	2569
OY	2270	AAATGCTACACACACAGCTCTGTGTGAATPACAAAGACGAGAAATGTCAAGAACTGAGCT	2339
Db	2564	TTTGAATTCATATGAAACATTAAGAGTATATGCAATTTCAACAACCTTTGACAAAAACAGTTT	2623
OY	2330	ACCTTGAGAAAGCTTATAGAGTCTCTGTAATACATCTACATCTGCCACAATATATCTCG	2389
Db	2624	ATCTAAGAAATTTGATGAATTTTTTAACTAAACATATATATGTATATATTTATTTTGG	2683
OY	2390	TCTCTCACAGCAGCATATGAGAGAGATCTTTAAACAGTACAAAGATTAACCAAGAGAGG	2449
Db	2684	TATCAACTCTAGATATGAGACCAAAATTTATAGAGTATATATCTTAC ---TCCAGAG	2740
OY	2450	AGAGTAAACTGTCTCTTGTGATCCACTGAGACCTGCTGTTCAATATCCAGAACACATTC	2509
Db	2741	AAATATGATTTAAATCATGTGATCGATGATTTATTTATTTAAATTCGAAATTAACATATAC	2800
OY	2510	CCGTTATGTATTTATGTTGGAATAGCCTCAACAATTTCTCTCTCAACTGTTATGAGAGA	2569

Db	2801	CHGCATGATGATTCATTATATGATATATGAGAAATGATTTACAAACATCTCTTTTGAAT	2860
Qy	2570	TATATGAGAAGAGATGGCTGCAACCTGATATAACTCAAGACAGCAAGATTAAAGA	2629
Db	2861	TATATCAAAAGAAATGATTTATTTATTTACATAAACTAAAGAGAGAAATACATCAAAA	2920
Qy	2630	ACCTCTCTGAGAGAACTAA	2648
Db	2921	AATTATTAGAGAGCAAA	2939
RESULT 28			
PEP190G			
LOCUS	PEP190G	3518 bp	DNA
DEFINITION	Plasmodium falciparum surface antigen P190 gene fragment.	linear	INV 29-NOV-1987
ACCESSION	Y00087		
VERSION	Y00087.1	GI:9924	
KEYWORDS	p190 gene; surface antigen.		
SOURCE	Plasmodium falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 3518)		
TITLE	Certa, U.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (09-NOV-1987) Ulrich C., Hoffmann-La-Roche, Department		
AUTHORS	ZfE, Grenzacher Str. 124, 4002 Basel, Switzerland		
TITLE	2 (bases 1 to 3518)		
JOURNAL	Certa, U., Roemann, D., Matile, H. and Reber-Liske, R.		
REFERENCE	A naturally occurring gene encoding the major surface antigen		
AUTHORS	precursor p190 of plasmodium falciparum lacks tripeptide repeats		
TITLE	EMBO J. 6 (1987) in press		
JOURNAL	Location/Qualifiers		
FEATURES	1..3518		
SOURCE	/organism="Plasmodium falciparum"		
	/strain="Ro-33"		
	/db_xref="taxon:5833"		
	/clone="RO-33/1"		
	337..29518		
	/note="p190 surface precursor (AA 1-107) (3518 is 2nd base		
	in codon)"		
	/codon_start=1		
	/protein_id="CA68280.1"		
	/db_xref="GI:9925"		
	/db_xref="SWISS-PROT:P19598"		
	/translation="MKIIFFLCSELFPIINTQCVTHESYOELVKLEALEDAVLGYS		
	LFQEKWMLKDGANTQVAVPADAVSTQSAKNPAGAVPSTGASTKGAISPGANPS		
	DDSSDPAKSYADLKHRVONLFTIKELKYEPLDNLNHHMLTLDNTHGFEYLIDGEE		
	EINELVILKINYPEPLRLAKRANINADVANCYCPENLKIIRANLELDVLRKIVGRPLDE		
	IKDVGKMEVDYIKRKNKTTIANINELIGSKRTDIONKNADEBKRLKYOAVDPLTYI		
	NKQJQEAHNLISVLEKRIIDTLKKNEINIKLEEDIDKIRIDAEKFTGVNQLSLERK		
	ESRREKIRELAKITIKENIDLFDPLELEYLEKKRVDPKSDPPKSVQIPVAV		
	PYNGYIVPLPLTDIHNSLADNDKNSGDIEMNPTKEINELKITDNKERKFIINNI		
	KQJQJDEKKNINHTKEONKRLLEDYKSKYDELLKFEYMRNNNEKNDVVDKIFES		
	ARVAYNEKORYNNKFSSSNNSVYVNOVKIKKALSYLDYSLRGISRDKPHVYTLTKT		
	GLEADIKLTLEIKSEKRIELKFKGLTHSANAASLESDYVQJQVOKVLLIKIEDDL		
	KRIELFLMAQLKDSIHVPNIYRPNQNPPEYVILVLRKVDKLEFIPKVDMLKKEQ		
	AVLSITQPLVAASETTEDEGHSHTLTOSQETVEETETETVGHATTVYITLPKREK		
	KVYENSLIEKNSDNSQALTKTVYIKLDEFLTKYSICHKYLIVSNSSMDOKLLEVNI		
	TPNEENSLKCDRDLFNIONNIIPAMYSIVDSNNNDLOHLEFFLEOKEMEYLLKLEK		
	ENHITKLLIEPKQITGTSSTSSPENTYVNNVNAQSTHNSNODSNAESTYNNQNVAS		
	SGPVAVESHDPLTVLSINDLKGVISLNLGNATKVPNPLITSTEMERTEINLTKI		
	MIPIFNDDIKQFVKSNSKRYITGLTETQKMAINDKIKLKDITLQSLFDLYNNYKRLKLD		
	LFNKKKEIGQDKQKIKKLTLLLEQLEKRLNS"		
BASE COUNT	1570 a	430 c	437 g 1081 t
ORIGIN			
Query Match	14.4%	Score 711.8;	DB 3; Length 3518;
Best Local Similarity	56.8%;	Pred. No. 1.6e-145;	
Matches 1511; Conservative	0;	Mismatches 1072;	Indels 76; Gaps 8;
10 ATGAATCATTTTCTCTCTGTCATTCATTTCTGTTTTTATCATCAATCATCAGCGCTG	69		

Db 337 ATGAAGATCATATCTTTTATGTTTCATTTCTTTTTTTTATTATAAATACACAAATGTGA 396
 QY 70 ACCGAGATCCTATCAGAGCTGTTAGAAACCTGAAAGCTTGGAAATGCCCTT 129
 Db 397 ACACATGAAAGTTATCAAACTGTCAAAAACATAGAAAGCTTTAGAAATGAGATTTG 456
 QY 130 ACCGATACAGCCCTGTCAGAGAGAGAAATGCTGAATGAAGAGAGAGATGTCACG 189
 Db 457 ACAGGTTATAGTTATTTCAAAAGGAAAAATGATTTAAAGATGAGACAAATACTCA 516
 QY 190 GCCGTTACACACACACACCCGCTTAAAGGCTGTGCTAGCGGTGCTCCGGTGG 249
 Db 517 GTTGTGCAAGCCCTGCAATGCTGTAAAGTACTCAAAAGTCTCAAAATCTCCAGTGTCT 576
 QY 250 TCTGTGCGCTCTGCGGCTCCGCTCCGCGGAGGCTGGCATCAGTGGCTGAGTG 309
 Db 577 ACAGTACCTTCAGGTTACTGCAAGTACTAAAGGCTGTATGAATCTCCAGGTGTGCA --- 633
 QY 310 GCAGCGCGGCTTCGCGGAGACAGTCGAAACCAATCATCTGACACACTGACGATTC 369
 Db 634 -----AATCCTTCAGATGATTCAGTGTATCA 660
 QY 370 GACGCCAGTCTTACGCCGACCTCAAGCACCGAGTGAGAACTATCTCCTCACTATCAG 429
 Db 661 GATGCTAAATCTTACGCTGATTTAAACATAGAGTTCAAAATTAATTAATTAATTA 720
 QY 430 GAGCTGAAGTACCCAGTGTTCGACCTCAATCATGCTGTGACAGTGTGATAC 489
 Db 721 GAATCCTAAATATCCGAACTCTTGTGATTTAACCAATCATATGTTAATCTTGTGTAT 780
 QY 490 ATTCAATGCTCAATATCTGATGATGAGCTTACGAGAGATCAATGACCTCTGTACAA 549
 Db 781 ATTCAATGCTCAATATTAATTAATGATGATGAGAAATTAATGAATTAATTAATTA 840
 QY 550 TTGAATTTCTACTGACTCTGAGGCGCAACCTGATGAGCTTTGGCCAACTACTAT 609
 Db 841 TTAACCTTTATTTGATTTATTAAGAGCAAAATTAATGATGATGATGATGATGAT 900
 QY 610 TGTCAATTCATCAATTTGAAGATCAGAGCAACGAGTGTGACCTATTTGAAGATTTG 669
 Db 901 TGTCAATTCATCTTCAATTTAAATTCGTGCAAAATGATGATGATGATGATGATGAT 960
 QY 670 GTCTTGGATATGCAAGCTCTCGACAAATCAAGACAAATGTGGGAAAGAT 729
 Db 961 GTGTGCGATATAGAAAACCATTAAGCTTATTAAGATATAGTAAAGAAAATGAGAT 1020
 QY 730 TTTATTTAAAAAATTAAGAACCATCGAGACATTAACGAGCTGTCAAGAAATCCAA 789
 Db 1021 TACATTTAAAAAATTAACCAACATAGCAAAATTAATGAATTAATGAAGAAATGAT 1080
 QY 790 AAGACATAGACAAAATTAAGAAATGCAACCAAGGAGGAAAGAAAGATTTGATCCAG 849
 Db 1081 AAAAAATTTGATCAAAATTAAGATGCAATTAAGAAAGGAAAAAATTAATTAACAA 1140
 QY 850 GCCCATACGACCTGCTCATATTAACAACAGCTTGAAGAACCCATCACTCATACAG 909
 Db 1141 GGTCAATATGATCTTTTATTTACATAAACAATTAACAAGACACATATATTAATTAAC 1200
 QY 910 GTACTGAGAGAGCGCATAGACACCTCAAGAGAAATGAATAATATCAAGAACTGCTGAC 969
 Db 1201 GTTTTAAAAAAGCTATGACCTTTAAAAAATTAAGAAACATTAAGAAATTAATGAA 1260
 QY 970 AAGATTAATGAATTAAGAAATCT--CCGCGAGCCAACTCTGGGAAACCCCTACACAG 1027
 Db 1261 GATTAATGAATTAATTAATTAAGATGCGGAAAAACCATGCTGAGATGAACCAAAATCTC 1320
 QY 1028 TGCCTGAGAGAGAGAGAGATGAGAGACAGAGAAAGAGATCAAGAGATCGCAAAA 1087
 Db 1321 TCCCTGAGATTTAGAAAAAGTCGAG--ACAGAGAAAAAATTAAGAAATTTGCCAAA 1378
 QY 1088 CCAATTAAGTTCACATAGATTTCTCTTTTACTGATCCCTTGAAGTGAATCTACTGA 1147
 Db 1379 CTATTAAATTTAAACATGATAGTTATTTACTGATCCACTGTAATTAATTAATTA 1438

QY 1148 GAGAGAGATTAAGATTAATGATACATCTCCGCCAAAGTCGA-----GACAAAGAAAT 1198
 Db 1439 GAGAAAAAATTAATAAGTTGATGTACACCTTAATACACAGATCTTACGAAATCTGTT 1498
 QY 1199 CAACCGAACCTTAATGAATATCCCAATGTGTGACGTACCTCTGTCTTATTAACGATATCA 1258
 Db 1499 AATATCAAAAGTCTTATCCAAATGTATGTATATCTCTTACACAGTACATGATATTC 1558
 QY 1259 ACAC-----CGCTTCACAGAGCTCAATATGCTTGGTGTACTGATTAACCCCTTG 1309
 Db 1559 ATATATTCATTAAGTGTGAGATTAATGAATTAATCAATGATGTGATTAATGAATCCCA 1618
 QY 1310 ATATATGAGAAAGACCTCTTAAGATATATGACACAGACATGAGAGAAAGATTTATCA 1369
 Db 1619 CTAAAGAAAAAATTAATGAATAATTAATTAACGATTAATTAAGAAACAAAAATTTATTA 1678
 QY 1370 ACGAAATCAAGAGAAATCAAAATTTGAGAGAGAAATTTGAGAGTGAACAGAAAGTT 1429
 Db 1679 ATATATTAATAAAGAAATTTGATTTAGAGAAAAAACAATTAATCAACAAAAAGAA 1738
 QY 1430 ACAGAGACCGCAGCAAAAGTCTTAACGATATCTTAAGAGATGAAAGTCTGACAG 1489
 Db 1739 ATAAAAAATTTACTTGAAGATTTATGAAGATCAAAAAAGGATTAATGAAGAAATTA 1798
 QY 1490 AGATCTATGTTCCAAATTCACAAATACATGACCTGACCAACTCGAGAAATGATG 1549
 Db 1799 AATTTATGAATTAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1858
 QY 1550 GAAACGATCTCTTACAAAGTGAAGAACTGACACACATATATACCTTGTGATCTATG 1609
 Db 1859 GTCCAGATATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1918
 QY 1610 AGAATTTCAAGCTATATCTTGAAGATCAACCAACTCTTAAGTATATGAGAGATAT 1669
 Db 1919 ATATATCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1978
 QY 1670 CTCTGCGGAACATTTGTTGAGAGAACTAAGATTTACAAAGTATCATTAAGTAA 1729
 Db 1979 CTTTAAGAAAGAAATTTCTGAAAAAGATTTTAATCATTAATTAATTAATTTGAAAC 2038
 QY 1730 TCGAAAAAGATCGAGACCTGTTGAGAACATTAAGAGATTAAGAGATGAAAGACCTTTG 1789
 Db 2039 TCGAAGCTATATTAATAAATTAACAGAAATTAAGAGATGATGATGATGATGATGAT 2098
 QY 1790 AGAAGAAATTTCAAAAGAGCAAAATTAACAGATGAGAAATCTCGAGGCTCCGATA 1849
 Db 2099 AAAAAATTTTAAAGGATTAACATCACTTACAGCAATGCTTC--CTTAAGATTTGATA 2155
 QY 1850 TTGTTAAAGTCCAAAGTGAAGAGTCTCTCATGAAACAAATTTGAATCAACAAAGAA 1909
 Db 2156 TTGTTAAATTTACAGTACAAAAAGTTTATTAATTAATAAATAATGAAGCTTAAGAA 2215
 QY 1910 CTCACCTCATTTGAGAAAGCTGAGATTAACATTAATTAATTAATTAATTAATTAATTA 1969
 Db 2216 TGAATTAATTTTAAAAATTAAGCACTAAAGATGATGATGATGATGATGATGATGAT 2275
 QY 1970 AGCAGAGATTAAGAGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2029
 Db 2276 AACCAACAAATTAACCGAGAACCATTAATTAATTAATTAATTAATTAATTAATTAATTA 2335
 QY 2030 TGAAGTGTTCATGCCCAAGTCGAGAGCTGTCAACGAGAGAGAAAGAAATTAATA 2089
 Db 2336 TAAAGAAATTTATACCAAAAGTAAAGACATGTTAAAGAAAGACAAAGCTGTCTTCA 2395
 QY 2090 CTGAAGACAGTCTGATTAATCTCGAGCTTCAACAGAAAGAGATTAACCGACAGCTTA 2149
 Db 2396 GTATTACAAACCTTAATGTTG-----CAGCAAGCGAAACACTGAAGATGAGG 2443
 QY 2150 CCAACAGGCCGAGACAAAGCGGTTGAGCTCTGCAAGCGGATTAACGTCAGCTCAAG 2209
 Db 2444 GTCACTCCACACACATTAATTCCTCAATGAGGAGAAACAGAAAGTAAACAGAAACCAAG 2503

Db 781 ACTACAGTATGCTTTTAAAAATCTGCTAAAAACCGTTTATCTAATTACTTGATGA 840
Oy 4567 AACCTGACGGGATGCTGAACATCTCCAGCAACCAATGGGTGAAGAAACAGCCCCAG 4626
Db 841 AACTTGCAAGATATGTTTAAACATTTTACACACCAATGGGTGAAGAAACAGTCCACAA 900
Oy 4627 AATAGCGGCTGTTTGAAGCATGAGCAGCCGGAAGAGTGCAGAGTCTCCGTAAGTAC 4686
Db 901 AATTCTGATGTTTACAGATTTAGATGAAGAGAAAGATGAATGAATGTTTATTAATTAC 960
Oy 4687 AAACAGAAAGGATAGTACGCTGGAGAACCAAAACCTTACTGCAATGAAGAAATGCGC 4746
Db 961 AAACAGAAAGGATGAATGAATGTTGAAATCCAAATCTTACTGTAACCAATAATATGCT 1020
Oy 4747 GGGTGTGACGGCATGCTAATATGACCCGAGAAACAGCGGCTTACAGGAAAGAAATC 4806
Db 1021 GGATGTGATGACAGATGCCAAATGTACCGAAGAAATGACGAGCAAGAAAGAAATC 1080
Oy 4807 ACATGCGAGTGTACTAAGCCGACATCTCCATCTTTCAGCGAGATTTTTCCTCAGC 4866
Db 1081 ACATGGAATGTACTAAGCCGATCTTATCCACTTTTGTGATGTTTTCGCAATGTC 1140
Oy 4867 TCTAATTTCTCGGATCTCTCTCTGCTGATCTCTGATCTGATCTGATCTGATCTGATC 4926
Db 1141 TCTAATCTCTTGAATATCATCTTATTAATCTCATCTGTTAATATATACAGTTTCATT 1200
Oy 4927 TAA 4929
Db 1201 TAA 1203

RESULT 30
AF325919 1131 bp mRNA linear INV 13-FEB-2001
LOCUS Plasmidium falciparum merozoite surface protein 1 (MSP-1) mRNA,
DEFINITION partial cds.
ACCESSION AF325919
VERSION AF325919.1 GI:12751398
KEYWORDS
SOURCE Plasmidium falciparum.
ORGANISM Plasmidium falciparum.
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS L.I.X.R., Goel,V.K., Liu,S.C., Chisholm,A.H. and Oh,S.S.
TITLE 42 kDa subfragment of MSP-1 gene of Plasmodium falciparum FCB-1 isolate
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1131)
AUTHORS L.I.X.R., Goel,V.K., Liu,S.C., Chisholm,A.H. and Oh,S.S.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2000) Section of Hematology/Oncology Research,
St. Elizabeth's Medical Center, Tufts University School of
Medicine, 736 Cambridge Street, Boston, MA 02135, USA

FEATURES
source
1. .1131
/organism="Plasmodium falciparum"
/isolate="FCB-1"
/db_xref="taxon:5833"
/country="Colombia"
1. .1131
/gene="MSP-1"
/gene="MSP-1"
/gene="MSP-1"
/codon_start=1
/product="merozoite surface protein 1"
/protein_id="AAK07641.1"
/db_xref="GI:12751399"
/translation="AATPSVIDNLSIKENIEYVLYLKLPLAGVRSLLKOLENNVMT
NVNKKDILNSRENKRNKLVESDLIPYKOLSSNVVKKPSEFLNKKRDKFLSY
NYIKSDIDTINPANDVILGYKILSEKKYKSDLSIKYINDKQENKYLPLNNIE
LYKTVNKRIDLPLYHLKALVNTYTESNVEVKIKELNLTATIDKDLKRNKPNFQ
IADLSDYNHNNLLTKFLSTGMEFENLAKTVLSMLDGNLQGLNLSIHOHCVKQCPQ
NSGCFRHLDERRECKLLNFKOEGDKCVENPNPCNENNGCDADACTBERDSNGK

BASE COUNT 467 a 136 c 154 g 374 t
ORIGIN
Query Match 13.3%; Score 655.8; DB 3; Length 1131;
Best Local Similarity 73.7%; Pred. No. 3e-133;
Matches 834; Conservative 0; Mismatches 297; Indels 0; Gaps 0;
Oy 3799 GCTGTACATCTCTCCGTGATGATTAACATTTCTGCCAAATCGAAGACGATACGAATG 3858
Db 1 GCAGTACCTCTCCGTGATGATTAACATTTCTGCCAAATCGAAGACGATACGAATG 60
Oy 3859 CTCATCTGCAACCTCTGGAGCGCTTATGATCTCTCAAGAAAGAGCGGATTAAC 3918
Db 61 TTATATTTAAACCTTTACGAGGTGTTTATAGAGTTTAAAGAAATAGAAATTAAC 120
Oy 3919 GTGATGACCTTCAATGTCACGTAAGACATTTCTCAAGCGCTTTAATAGAGAA 3978
Db 121 GTTATGACATTTAATGTTAATGTTAAGATATTTTAAATTCAGGATTTAATAGCTGAA 180
Oy 3979 AATTTCAGAGCTCTTGAGAGCGCTTATGATTTCTTATTAAGACCTGACCTCTTAAC 4038
Db 181 AATTTCAGAGCTCTTGAGAGCGCTTATGATTTCTTATTAAGACCTGACCTCTTAAC 240
Oy 4039 TACGTTGCAAGGACCATTAACATGCTCAATTAAGAGAGGATTAATTTCTGCT 4098
Db 241 TATGTTGCAAGGACCATTAACATGCTCAATTAAGAGAGGATTAATTTCTGCT 300
Oy 4099 AGTTTCAACTATATCAAGACTCTCATCGACACCGATATCAATTTGCTATGCTG 4158
Db 301 AGTTTCAACTATATTAAGACTCTCATCGACATTAAGATTAATTTGCTATGCTG 360
Oy 4159 GGGTATTACAGATCTCTGAGCGGAAATTAACAGCTGACCTTACTTATTAAGAT 4218
Db 361 GGGTATTACAGATCTCTGAGCGGAAATTAACAGCTGACCTTACTTATTAAGAT 420
Oy 4219 ATCAACGATTAAGAGGAGAGATGAATGAATTAATTCCTCTGATTAACATTCGAAC 4278
Db 421 ATCAACGATTAAGAGGAGAGATGAATGAATTAATTCCTCTGATTAACATTCGAAC 480
Oy 4279 CTGTACAGACAGTGAAGCAAAATCGACTCTTCTGATTAACATTCGAACAGCTC 4338
Db 481 TTATATTAAGAGGATTAATGAATTAATTCCTCTGATTAACATTCGAACAGCTC 540
Oy 4339 CTCACATTAATCTTACAGAGAGCAATGTGGAATTAATTAAGAGAGCTGACCTGCTC 4398
Db 541 CTCACATTAATCTTACAGAGAGCAATGTGGAATTAATTAAGAGAGCTGACCTGCTC 600
Oy 4399 AAACCAATCCAGACAGCTGCGAGATTTCAAGAAATTAACATTTCTGCGAATTCGA 4458
Db 601 AAACCAATCCAGACAGCTGCGAGATTTCAAGAAATTAACATTTCTGCGAATTCGA 660
Oy 4459 GACCTGTCTACAGGATTAATTAACCAATCTCTGACCAAGTTCTGTCACAGCTGCTC 4518
Db 661 GATTTATCAACAGATTAATTAACCAATCTCTGACCAAGTTCTGTCACAGCTGCTC 720
Oy 4519 GTTCTGAAAACCTGCGCAAAACAGTCTGACCAATCTCTGACCAAGCTGCAAGGCTC 4578
Db 721 GTTCTGAAAACCTGCGCAAAACAGTCTGACCAATCTCTGACCAAGCTGCAAGGCTC 780
Oy 4579 ATGCTGACATCTCCAGACCAATGCGGAGAAACAGTGGCCCGCAGAAATGCGGCTG 4638
Db 781 ATGCTGACATCTCCAGACCAATGCGGAGAAACAGTGGCCCGCAGAAATGCGGCTG 840
Oy 4639 TTCAGCATCTGACAGAGCGGAGAGTGAAGTGTCTCTCACTACATCAACAGAGTGA 4698
Db 841 TTCAGCATCTGACAGAGCGGAGAGTGAAGTGTCTCTCACTACATCAACAGAGTGA 900
Oy 4699 GATTAAGTGTGAGAGAGAGCAACCTTACCTGATGAAGCAATAGCGGGGTGTACGCC 4758
Db 901 GATTAAGTGTGAGAGAGAGCAACCTTACCTGATGAAGCAATAGCGGGGTGTACGCC 960
Oy 4759 GATGCTAATGACAGGAGAAAGACAGCGGCTTACGGAAGAAATACATGCGAGTGT 4818


```

Db 961 GATGCAAAATGTCACCAAGAGATTCAGTAGCAGCAAGAAAGAAATACATGATGATG 1020
Qy 4819 ACTAACCCGACACTCTATCCACTCTCTGACGCGGATTTTTCCTCCAGCTATTTCTG 4878
Db 1021 ACTAACCGATGATTTTATCCACTCTCTGACGATGATTTTCTGCGCTCTCTA 1080
Qy 4879 GGCATCTCTCTCTGATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 4929
Db 1081 GGAATATCATCTTATTTATATCATGATGATTTATTTATGAGTTTCACTTTAA 1131

RESULT 31
PFASURFPRO 1065 bp DNA linear INV 26-JUL-1993
LOCUS Plasmodium falciparum merozoite surface protein (MSP-1) gene
DEFINITION
ACCESSION L20092
VERSION L20092.1 GI:309745
KEYWORDS merozoite surface protein.
SOURCE Plasmodium falciparum (strain Vietnam Oak Knoll) blood stage DNA.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1065)
AUTHORS Louis-Willeman, V., Shi, Y., Collins, W. and Lal, A.
TITLE Primary amino acid sequence of the carboxyl-terminal region of the merozoite surface protein (MSP-1) of Plasmodium falciparum Vietnam Oak Knoll (FVO) strain
JOURNAL Unpublished (1993)
FEATURES
    source          location/Qualifiers
                    1..1065
                    /organism="Plasmodium falciparum"
                    /strain="Vietnam Oak Knoll"
                    /db_xref="taxon:5833"
                    /dev_stage="blood stage"

BASE COUNT 446 a 125 c 150 g 344 t
ORIGIN
Query Match 12.6%; Score 623.6; DB 3; Length 1065;
Best Local Similarity 74.2%; Pred. No. 3.5e-126;
Matches 788; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

Qy 3799 GCTGTACACCTCTGCTGATGATGATTAACATCTCTCCAAATTCAGAAACGATACGAAGT 3858
Db 1 GCGATACCTCTCTGCTGATGATGATTAACATCTCTCTAAATTTGAATGATGAGTT 60
Qy 3859 CTTCTATCTGAAACCTCTGCGAGCGCTCTATAGCTCTCTCAAGAAACAGTGGAGAAATAC 3918
Db 61 TTATATTAAACCTTAGAGAGTGTATAGAGTTTAAAGAAACATTTAGAAATATAC 120
Qy 3919 GTGATGACCTTCAATGTCAACGTGAAGACATTTCTGAAACGCCGTTTAAATGAAGAGAA 3978
Db 121 GTATATGACATTTTAAATGTTAATGTTAAGATATTTTAAATTCAGATTTAATAACGGAA 180
Qy 3979 AATTTAAGAAACCTTTGGAGAGCGACTGATTCCTATTAAGACCTGACCTCTCTAAC 4038
Db 181 AATTTAAGAAACCTTTAGATCAGATTTAATTCATTAAGATTTAATCAATCAAGTAAT 240
Qy 4039 TACGTTGCAAGAGACCATTCAGATTTCTCTCAATTAAGAGAGGGGTTAATTTCTGCT 4098
Db 241 TATGTTGTCAGAGATCCATTAATTTCTTAATTAAGAAAGAGATTAATTTCTTAAGC 300
Qy 4099 AGTTAACAATATCAAGAGACTCCATCGACAGCATATCAATTTCCGTAAATGATGCTG 4158
Db 301 AGTTAATAATATTAATTAAGAGATTCATTAAGATTAATTTTTCGAATATGATGTTCT 360
Qy 4159 GGGTATTAACAAGATCTGAGCGAAATATCAAGTTCGACCTGATTTAAAGATAT 4218
Db 361 GGAATATTAATAATTAATTCGAAATATTAATTAATTAATTAATTAATTAATTAAT 420
Qy 4219 ATCAAGATTAAGAGAGCGAGATTAAGAAATATCTGCTCTCGAATTAATCAATCGAAAC 4278
Db 421 ATCAAGCAAAACAGGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480

```

```

Qy 4279 CTGTACAGACACTGAACGCAAAATGCACTCTCTGTAATTCACCTGGAGCCAGGTC 4338
Db 481 TTATATTAAGAACAGTATTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 540
Qy 4339 CTCACTATTAATTAATTAAGAGAGCAATGATGAGTTAAATTAAGAGAGCTGACCTAC 4398
Db 541 CTAAATTAATTAATTAATTAAGAGAGCAATGATGAGTTAAATTAAGAGAGCTGACCTAC 600
Qy 4399 AAACATCCCAAGACAGCTGGAGATTTTCAGAAAAATTAACATTTTCGTCGGAATGCA 4458
Db 601 AAACATTTCAAGCAAAATTTGCGAGATTTTAAAAAATTAACATTTTCGTCGGAATTTGCT 660
Qy 4459 GACCTGCTACCGATTAATTAACCAACAAATCTCTGACCAAGTTCTGTCACCTGCGATG 4518
Db 661 GATTTATCAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 720
Qy 4519 GTGTTGGAAGAACTCGCCCAAAACAGTCTGAGCAATCTGTCAGCGCAACCTGCAAGGC 4578
Db 721 GTTTTGAAGAAATCTCTTAACCGTTTATCTAATTTACTGATGATGAAACCTGCAAGGT 780
Qy 4579 ATGCTGACATCTCTCCAGCACCAGTCTGTAAGAAACAGTGGCCAGATTAACGCGCTGT 4638
Db 781 ATGTTAAACATTTTACACAGCAACATGCGTAAATAAATGTCACAAATTTCTGATGT 840
Qy 4639 TTCAGCATCTGACGAGCGCGAAGAGTGCAGTGTCTCTGACTACAAACAAAGAGA 4698
Db 841 TTCAGCATTTAGATGATGAAGAGAAATGTAAATTTAATTAACAAAGAGAGT 900
Qy 4699 GATAAGTGGTGGAGAAACCAACCTACTGCAATGAAGAAACAGTGGCGGTGACGCC 4758
Db 901 GATTAATGTTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
Qy 4759 GATGCTAAATGACAGCAGAGAGAGACAGCGCTCTAAGGAAAGAAATACATGCGAGTGT 4818
Db 961 GATGCAAAATGATTCAGAGAGAGATTCAGTGAAGCAAGGAAAGAAATACATGATGATGT 1020
Qy 4819 ACTAACCCGACACTCTATCCACTCTCTGACGCGGATTTTTCG 4860
Db 1021 ACTAACCTGATCTTATTCACCTTTTCGATGATTTTTCG 1062

RESULT 32
PFMSA1AC 999 bp DNA linear INV 22-MAY-2002
LOCUS Plasmodium falciparum clone 808 major merozoite surface antigen
DEFINITION (p190) gene, partial cds.
ACCESSION M77715
VERSION M77715.2 GI:21040168
KEYWORDS T-cell epitope, major merozoite surface antigen, tripeptide repeat.
SOURCE Plasmodium falciparum.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 999)
AUTHORS Jongwutives, S., Tanabe, K., Nakazawa, S., Yanagi, T. and Kanbara, H.
TITLE Sequence variation in the tripeptide repeats and T cell epitopes in p190 (MSP-1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
COMMENT
    JOURNAL 92228018
    MEDLINE 1373473
    PUBMED 2 (bases 1 to 999)
    REFERENCE 2 (bases 1 to 999)
    TITLE Jongwutives, S.
    AUTHORS Direct Submission.
    JOURNAL Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn University, Rama4, Pathumwan, Bangkok 10330, Thailand
    On May 21, 2002 this sequence version replaced gi:309698.
FEATURES
    source          location/Qualifiers
                    1..999
                    /organism="Plasmodium falciparum"
                    /db_xref="taxon:5833"
                    /map="9"
                    /clone="808"
                    /sex="asexual"

```


Db 1 GCTTTAGAGATGACAGTATTGACAGGTTTAAAGTTTATTTCAAAAAGAAAAATGATTTA 60
 QY 169 AATGAAGGACGAGTGGCGCGGCTTACAAACGACACCGCGTCTTAAAGGCTGTG 228
 Db 61 AATGAAGGACGAGTGGCGCGGCTTACAAACGACACCGCGTCTTAAAGGCTGTG 120
 QY 229 GCTAGCGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGTTCCGTCGCTCCGGCGCAGC 288
 Db 121 GCTTACAGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGTTCCGTCGCTCCGGCGCAGC 180
 QY 289 GTGGCATCAGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGTTCCGTCGCTCCGGCGCAGC 348
 Db 181 GTTGTCTCAGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGTTCCGTCGCTCCGGCGCAGC 240
 QY 349 TCTGCAACTCTAGCGATTCGACGCGCAAGTCTTCCAGCCGCTCAAGCAACGAGTGA 408
 Db 241 TCAGATTAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 409 AACTATCTCTCAGTATCAAGAGAGTGAAGTACCACAGTGTTCGACCTCCTATCAT 468
 Db 301 AATTAATCTCTCAGTATCAAGAGAGTGAAGTACCACAGTGTTCGACCTCCTATCAT 360
 QY 469 ATGCTGACACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 528
 Db 361 ATGCTGACACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 QY 529 ATCAATGACACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 588
 Db 421 ATTAATGACACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 QY 589 GACGTTTGGCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 648
 Db 481 GAT 540
 QY 649 TTGAGCTATTTGAAGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 708
 Db 541 TTGAGCTATTTGAAGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 QY 709 AATGTGGGAAAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 768
 Db 601 AATGTGGGAAAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 QY 769 GACCTGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGAT 828
 Db 661 GACCTGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGAT 720
 QY 829 GAAAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 888
 Db 721 GAAAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 QY 889 GAAAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 948
 Db 781 GAAAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 QY 949 AATATCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1008
 Db 841 AATATCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 QY 1009 GGAAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1068
 Db 901 GGAAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 QY 1069 ATCAAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107
 Db 961 ATCAAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 999

RESULT 34 PFMSA1AM 999 bp DNA linear INV 22-MAY-2002
 LOCUS PFMSA1AM
 DEFINITION Plasmodium falciparum clone 837 major merozoite surface antigen
 ACCESSION M7725

VERSION M7725.2 GI:21040188
 KEYWORDS t-cell epitope; major merozoite surface antigen; tripeptide repeat.
 SOURCE Plasmodium falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 999)
 AUTHORS Jongwutiwes, S., Tanabe, K., Nakazawa, S., Yanagi, T. and Kanbara, H.
 TITLE Sequence variation in the tripeptide repeats and T cell epitopes in
 P190 (MSA-1) of Plasmodium falciparum from field isolates
 JOURNAL Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
 MEDLINE 92228018
 PUBMED 1373473
 REFERENCE 2 (bases 1 to 999)
 AUTHORS Jongwutiwes, S.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn
 University, Rama4, Pathumwan, Bangkok 10330, Thailand
 COMMENT On May 21, 2002 this sequence version replaced gi:309716.
 FEATURES
 source
 1..999
 /location="Qualifiers
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /map="9"
 /clone="837"
 /sex="asexual"
 /dev_stage="erythrocytic"
 <1..>999
 /gene="p190"
 <1..>999
 /gene="p190"
 /codon_start=1
 /product="major merozoite surface antigen"
 /protein_id="AAA29672.2"
 /db_xref="GI:21040189"
 /translation="ALEDAVLTGYSLFQEKMYLNBSGTAVTSTPSKGSVASG
 SGGSVASGGSVASGVSAGSVASGSGRSRTNSDSDASVADLKRVRNY
 LTIKRLKYPQLEPLFNHMLTCDNHGKRYLLDGEYELNELLKYNFEDLLRATLN
 DVCANDYCOIPENLKRANBLDYLKLVYGRPLDNIQDNQKMDYIKRKNKTLEN
 IMELIEESKRTIDKNRNAREEERKRLYQYDLSLYNQLEAHNLISVLEKRDITL
 KKNENIKELDLKINEIKNPSPASGNTPTLLDKNKIEHEKEIKETIKFNID"
 BASE COUNT 418 a 127 c 159 g 295 t
 ORIGIN
 Query Match 11:58; Score 570.2; DB 3; Length 999;
 Best Local Similarity 73.2%; Pred. No. 1.9e-114;
 Matches 731; Conservative 0; Mismatch 268; Indels 0; Gaps 0;
 QY 109 GCTTTGAGAGATGCGGCTTACCGGATGACGCTGTCCAGAGAGAGATGCTGTG 168
 Db 1 GCTTTGAGAGATGCGGCTTACCGGATGACGCTGTTCAAAAGAAAAATGATTTA 60
 QY 169 AATGAAGGACGAGTGGCGCGGCTTACAAACGACACCGCGTCTTAAAGGCTGTG 228
 Db 61 AATGAAGGACGAGTGGCGCGGCTTACAAACGACACCGCGTCTTAAAGGCTGTG 120
 QY 229 GCTAGCGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGTTCCGTCGCTCCGGCGCAGC 288
 Db 121 GCTTACAGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGTTCCGTCGCTCCGGCGCAGC 180
 QY 289 GTGGCATCAGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGTTCCGTCGCTCCGGCGCAGC 348
 Db 181 GTTGTCTCAGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGTTCCGTCGCTCCGGCGCAGC 240
 QY 349 TCTGCAACTCTAGCGATTCGACGCGCAAGTCTTCCAGCCGCTCAAGCAACGAGTGA 408
 Db 241 TCAGATTAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 409 AACTATCTCTCAGTATCAAGAGAGTGAAGTACCACAGTGTTCGACCTCCTATCAT 468
 Db 301 AATTAATCTCTCAGTATCAAGAGAGTGAAGTACCACAGTGTTCGACCTCCTATCAT 360
 QY 469 ATGCTGACACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 528

Db 361 ATGTTAACTTTGTGTGATTAATATTCATGTTCAAAATATTTAATGTGATGATGAAGAA 420
QY 529 ATCAATGACCTCCGTACACACTTGAATTTCTACTTCGACTTGCTAAGGCCAACTGAAT 588
Db 421 ATTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
QY 589 GACCTTTGGCCCAATGACTATTTGCAATTCATTTCAATTTGAAGTACAGACCAAGAG 648
Db 481 GATGTATGTGCTAATGATTAATTTGCAATTTCAATTTCAATTTCAATTTCAATTTCA 540
QY 649 TTGACGATTTGAAGAGTTGGTCTCGATATGCAAGCCCTTCGACACATCATGAGAC 708
Db 541 TTAGACGACTTAATAAACTGTGTGCGATATGAAACCATTAGACAAATATTAATGAT 600
QY 709 AATGTGGAAAGATGAAAGATTAATTAATAAGATAAGAACCATTCGAGACATTTAC 768
Db 601 AATGTAGAAAAGATGAAAGATTAATTAATAAGATAAGAACCATTCGAGACATTTAC 660
QY 769 GAGCTGATCGAAGATCCAAAAGACCATAGACAAAATAAGATGCAACCAAGAGAGAA 828
Db 661 GAATTAATTTGAAGAGATTAAGAAAACATTTGATTAATAATGACATTAAGAGAGAA 720
QY 829 GAAAAGAGAGATTTGACACGCTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 888
Db 721 GAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
QY 889 GAAGCCCATACCTCATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
Db 781 GAACCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
QY 949 AATATCAAGAACTGCTGACACAGATTAATTAAGAAATTAAGAAATTAAGAAATTAAG 1008
Db 841 AACTTAAGAGAACTGCTGACACAGATTAATTAAGAAATTAAGAAATTAAGAAATTAAG 900
QY 1009 GGGACACCCCTTAACACGCTGCTGACACAGACAAAGATAGAGAGACAGAGAGAG 1068
Db 901 GGAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
QY 1069 ATCAAGAGATCGCCAAACCATTAAGTTCAACATGAT 1107
Db 961 ATAAAGAAATTTGCCAAACATTAATTAATTAATTTGAT 999

RESULT 35
PFAMSA1AA
LOCUS
DEFINITION Plasmodium falciparum clone 806 major merozoite surface antigen
(p190) gene, partial cds.
ACCESSION M77713.2 GI:21040164
VERSION T-cell epitope: major merozoite surface antigen; tripeptide repeat.
KEYWORDS
SOURCE Plasmodium falciparum.
ORGANISM
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
AUTHORS Jongwutives,S., Tanabe,K., Nakazawa,S., Yanagi,T. and Kanbara,H.
TITLE Sequence variation in the tripeptide repeats and T cell epitopes in
p190 (MSP-1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
MEDLINE 9228018
PUBMED 1373473
REFERENCE 2 (bases 1 to 999)
AUTHORS Jongwutives,S.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn
University, Rama4, Pathumwan, Bangkok 10330, Thailand
REMARK
COMMENT On May 21, 2002 this sequence version replaced gi:309694.
FEATURES
location/Qualifiers
1..999
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/map="9"

/clone="806"
/sex="asexual"
/dev_stage="erythrocytic"
<1..>999
/gene="p190"
<1..>999
/gene="p190"
/codon_start=1
/product="major merozoite surface antigen"
/protein_id="AA29660.2"
/translation="ALSDAVLTGSLFQEKMYLNEGTSGTAVTTSPGSGSVASG
SGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSVASG
LRTIKELKYPOLDFPHMLTIDNDNIFGVEYLIDYVEINELIKLNFYEDLLRAKIN
DYCANDCOQIPENIKITANELDYLKIVFRRPLDINRKNVKNMEYIRKKNKTEN
INELIESKRTIDNKNKATSEERKTIQOYDUSTINKOLEHNHLSVEKRIDPL
KKNENIELDKLINEIKNPPANSNPPLLDKNKIEHEREKIEIAKTIFNID"
BASE COUNT 418 a 128 c 159 g 294 t
ORIGIN

Query Match 11.5% Score 568.6; DB 3; Length 999;
Best Local Similarity 73.18; Pred. No. 4.3e-114;
Matches 730; Conservative 0; Mismatches 289; Indels 0; Gaps 0;

QY 109 GCTTTGAAGATGCGCTCCTTACCGGATACAGCTGTTCCGAAGAGAAAGATGCTGTG 168
Db 1 GCTTTAGAGATGCGAGTATTTGACAGGTTATAGTTATTTCAAAAAGAAAATGATATTA 60
QY 169 AATGAGGAGAGATGCGACGCGCTTACAAACACACACCCGGTCTTAAAGGCTGTG 228
Db 61 AATGAGGAGAACACGACGACAGCTGTTCACACTGATACCTGGTTCAAAAGGTTCAAGTT 120
QY 229 GCTAGCGGTGCTCCGGGGTCTGTGCGCTGCGGGGTTCGGTCCGCGGGGCGACG 288
Db 121 GCTTAGGTTGTTAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 289 GTGACATGAGTGGCTCAGTGGCAAGCGCGGTTCCGGGAACAGTGAAGAACCAATCCA 348
Db 181 GTTGCTTAGTGGCTCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 349 TCTGACAACTGTAGGATTCGACGCCAAGTCTACGCCGACTCAGACCGAGTGA 408
Db 241 TCAGATTAATTCAGATGATTCAGATGCTTAACTTACGCTGATTTAAACACAGAGTACGA 300
QY 409 AACTATCTCCCTACATTAAGAGAGTGAAGTACACACAGTGTTCGACTCCTCAATCAT 468
Db 301 AATTAATCTGTTAACTATTAAGAACTCAATATCTCTCACTTTGATTTAACTATCAT 360
QY 469 ATGCTGACACTGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 528
Db 361 ATGTTAACTTTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420
QY 529 ATCAATGACCTCCGTACACACTTGAATTTCTACTTCGACTTGCTAAGGCCAACTGAAT 588
Db 421 ATTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
QY 589 GACCTTTGGCCCAATGACTATTTGCAATTCATTTCAATTTGAAGTACAGACCAAGAG 648
Db 481 GATGTATGTGCTAATGATTAATTTGCAATTTCAATTTCAATTTCAATTTCAATTTCA 540
QY 649 TTGACGATTTGAAGAGTTGGTCTCGATATGCAAGCCCTTCGACACATCATGAGAC 708
Db 541 TTAGACGACTTAATAAACTGTGTGCGATATGAAACCATTAGACAAATATTAATGAT 600
QY 709 AATGTGGAAAGATGAAAGATTAATTAATAAGATAAGAACCATTCGAGACATTTAC 768
Db 601 AATGTAGAAAAGATGAAAGATTAATTAATAAGATAAGAACCATTCGAGACATTTAC 660
QY 769 GAGCTGATCGAAGATCCAAAAGACCATAGACAAAATAAGATGCAACCAAGAGAGAA 828
Db 661 GAATTAATTTGAAGAGATTAAGAAAACATTTGATTAATAATGACATTAAGAGAGAA 720
QY 829 GAAAAGAGAGATTTGACACGCTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 888

Db 721 GAAAAAATAATATACCAAGCTCAATATGATCTTCTATTACATAAACAATTAGAA 780
Qy 889 GAAGCCATATACCTCATACCGCTACTGAGAACCGATAGACACCTCAAGAAGATGAA 948
Db 781 GAAACCATATATTAATTAAGCGCTTTAGAAAAAGTATTCACACTTAAAAAAATGAA 840
Qy 949 AATATCAAGAACTGCTCGACAGATTAATGAATTAAGAACTCTCCGACGCACTCT 1008
Db 841 AACATTAGGAATTAATTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTA 900
Qy 1009 GGGACACCCCTTAACACGCTGCTGACAGAACAGAGATAGAGACGAGAAAGAG 1068
Db 901 GGAATAATACCAATATCTCTCTGATAGAACAAAAAATCGAGAACACGAAAAAGAA 960
Qy 1069 ATCAAGAGATGCGCAAAACCATTAAGTTCACATAGAT 1107
Db 961 ATAAAGAAATTCGCAAAACTATTAATTAATTAATTCAT 999

RESULT 36
PFAMSAIAP 999 bp DNA linear INV 22-MAY-2002
LOCUS Plasmodium falciparum clone 822B major merozoite surface antigen
DEFINITION (p190) gene, partial cds.
ACCESSION M77718
VERSION M77718.2 GI:21040174
KEYWORDS T-cell epitope; major merozoite surface antigen; tripeptide repeat.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 999)
AUTHORS Jongwutives,S., Tanabe,K., Nakazawa,S., Yanagi,T. and Kanbara,H.
TITLE Sequence variation in the tripeptide repeats and T cell epitopes in
p190 (MSA-1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
MEDLINE 92228018
PUBMED 1373473
REFERENCE 2 (bases 1 to 999)
AUTHORS Jongwutives,S.
TITLE Direct Submission
JOURNAL Submitted (21-May-2002) Faculty of Medicine, Chulalongkorn
University, Rama4, Pathumwan, Bangkok 10330, Thailand
COMMENT On May 21, 2002 this sequence version replaced gi:309703.
FEATURES
source location/Qualifiers
1..999
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/map="9"
/clone="822B"
/sex="asexual"
/dev_stage="erythrocytic"
<1..>999
/gene="p190"
<1..>999
/gene="p190"
/codon_start=1
/product="major merozoite surface antigen"
/protein_id="AAA29665.2"
/db_xref="GI:21040175"
/translation="ALEDAVLNGYSLFOREKWLNEGTSIGTAVTSTPGSGSVASGG
SGGSYASGVSAGSVASGSGSVASGSRTPSDSDSKSYADLKHRYNY
LITIKELKPOLFDLTNHLTCDNHGKRYLIDGEEINELLYKLNFEFDLRAKLN
DYCANDYCOIPNLKIRANIELDLVLRKLYGKRPIDNIRKDNVGGKEDYTKRKRTTEN
INELIEESKRTIDKNNKNTKEBEKKLYOAOYDISYKNQLEAHNLISVLEKRTDL
KKNENIKELDLKINEIKNPNSGNTPTLLDKNKIEHEKEIKELAKTIKENID"

gene
CDS
BASE COUNT 418 a 128 c 159 g 294 t
ORIGIN

Query Match 11.5%; Score 568.6; DB 3; Length 999;
Best Local Similarity 73.1%; Pred. No. 4.3e-114;
Matches 730; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

Qy 109 GCTTTGAGAGATCCGCTCCTTACCGATACAGCCTGTTCAGAGAGAGAGATGCTG 168
Db 1 GCTTTGAGAGATCAGATTTGACAGGTTATGATTTATTTCAAGAAAAATGATTA 60
Qy 169 AATGAGGAGACGAGTGGACAGCGCTTACACACACACCCGGTCTTAAGGCTGTG 228
Db 61 AATGAGGAGAACGAGGAGACAGCTGTTACACATGATACCTGGTTCACAGGTT 120
Qy 229 GCTAGAGGTCGTCGCGGTGCTGTGGCCCTGGGGGGTCCGTCGCTCCGCGCAG 288
Db 121 GCTTCAGGTGCTTCAGGTGCTCAGTTCGTTCAAGTGGCTCAGTTCGTTAGGTT 180
Qy 289 GTGACATCAGTGGCTCAGTGGCAGCGGCGGCTTCGCGGACAGTGAAGAACATCA 348
Db 181 GTTGTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAG 240
Qy 349 TCTGACACTCTGACATTCGCCAGCCCAAGCTCTACCGGACCTCAAGCAGAGTGA 408
Db 241 TCAGATTAATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 300
Qy 409 AACTATCTCCTCCTATTCAGAGAGCTGAAGTACCAAGTGTTCGACCTCACTAATCA 468
Db 301 AATTTACTTGTATTAATTCAGAGAGCTCAATATCTCAACTCTTTGATTTAATTCAT 360
Qy 469 ATGCTGACACTGTGTATTAACATTCAGCTTCATTAATTCATGATGACGTTACGAAG 528
Db 361 ATGTTAATCTTGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
Qy 529 ATCAATGAACTCTGTACAGATTTCTACTTGACTGTCTAGAGGCCAATCGAT 588
Db 421 ATTAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
Qy 589 GAGCTTGGCGCCATGATGATTCATTAATTCATTAATTCATTAATTCATTAATTCAT 648
Db 481 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
Qy 649 TTGAGAGTATTAAGAGATTTGATGATGATGATGATGATGATGATGATGATGAT 708
Db 541 TTGAGAGTATTAAGAGATTTGATGATGATGATGATGATGATGATGATGATGAT 600
Qy 709 AATGCTGAGAAAGTGAAGATTTATTAATTAATTAATTAATTAATTAATTAATTA 768
Db 601 AATGCTGAGAAAGTGAAGATTTATTAATTAATTAATTAATTAATTAATTAATTA 660
Qy 769 GAGCTGATCGAAGATTCACAAAAGACCATGACAAAATTAAGATTCACAAAAGAGAA 828
Db 661 GATTTATTTGACAAAGTAAAGAAACATGATTAATTAATTAATTAATTAATTAAT 720
Qy 829 GAAAGAGAGAGTGTACAGCGGCCAGTACGACCTGTCTATTAACAAACAGTTGAA 888
Db 721 GAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
Qy 889 GAAGCCATATACCTCATACCGCTACTGAGAACCGATAGACACCTCAAGAAGATGAA 948
Db 781 GAAGCATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
Qy 949 AATATCAAGAACTGCTCGACAGATTAATGAATTAAGAACTCTCCGACGCACTCT 1008
Db 841 AACATTAGGAATTAATTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTA 900
Qy 1009 GGGACACCCCTTAACACGCTGCTGACAGAACAGAGATAGAGACGAGAAAGAG 1068
Db 901 GGAATAATACCAATATCTCTCTGATAGAACAAAAAATCGAGAACACGAAAAAGAA 960
Qy 1069 ATCAAGAGATGCGCAAAACCATTAAGTTCACATAGAT 1107
Db 961 ATAAAGAAATTCGCAAAACTATTAATTAATTAATTAATTCAT 999

RESULT 37
AF061143 1133 bp DNA linear INV 09-FEB-2001
LOCUS Plasmodium falciparum isolate HN2 merozoite surface protein 1
DEFINITION

TITLE Direct Submission
 JOURNAL Submitted (12-JUL-2000) Section of Hematology/Oncology Research, St Elizabeth's Medical Center, Tufts University School of Medicine, 736 Cambridge Street, Boston, MA 02135, USA

FEATURES

Location/Qualifiers

1..1080

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/clone="FCB-1"

<1..>1080

/gene="msp-1"

<1..>1080

/gene="msp-1"

/note="38 kDa portion of MSP-1; MSA-1"

/codon_start=1

/product="merozoite surface protein 1"

/protein_id="AF87595.1"

/db_xref="GI:9438195"

/translation="QDKPEVSANDPTSHSTNLNLSLKENILSLGKKNITVQELIQ
 KSENFYEKILKSDPTFYNESFTNFVKRADINSLNDESKRLKLEEDINKLKLQ
 SEDLYNKKRLERLEKDKKTVGKKMOIKRLTLKLEQSLNLPKHYLOHFSV
 FENKKEAEIETENTLENTKILKHYGLKYYGESSPLKTLSEESTIOTEDNVSL
 ENFVLSKLEKIKONLLEKKKLSYSSGLHLIARLEKVKKNKNTGNSPEVNTD
 VNNALESKRKLPEPTDYATVYSESGSTLEDSQPKRPASTHVGAESNTITTSQNVDD
 EVDVIVPIRSESEEDYDLDGQVLT"

BASE COUNT

510 a 118 c 147 g 305 t

ORIGIN

Query Match 11.1%; Score 547.8; DB 3; Length 1080;
 Best Local Similarity 69.2%; Pred No.1.6e-109;
 Matches 747; Conservative 0; Mismatches 332; Indels 0; Gaps 0;

2713 CAAGACAAACCCGAGAGTGAAGCGCTAAGACGACACCTCTCACTCGACCAACCTTAAATAC 2772
 1 CAGGATAAACCCGAGAGTGAAGCGCTAAGACGACACCTCTCACTCGACCAACCTTAAATAC 60
 2773 TCACGAAACCTTTGAGAACATCCGCTCTCGGCAAGAAATAGACATCTACCAAGA 2832
 61 AGTTAAATTTTGAAGAAATGATGAGTCTTGGAACCAAAATATATACCAAGAA 120
 2833 CTATTGACAGAAATCGCCGAGAACTCTACGAGAACTGTAAGACGACGACACA 2892
 121 TTAATATGCGCAAAAAGTGTGAAAACCTTTATGAAAACATATTAATAATAGTATACA 180
 2893 TTCTATTAACGAGAGCTTCACTACTCTGTAATCTTAAACCGAGATATCAACTCTCT 2952
 181 TTTTATTAATGATCTTTTACAAATTTTGAATTTTGAAGCTGATGATATTAATCATG 240
 2953 AACGATGATCTTAAGCTGAAGAGCTGGAAGAGACATCAATATAGCTGAAGAAACACG 3012
 241 AATGATGAATCAAAAAGGAAATTAAGAAGATATTAATAATTAATAAAACCTTTA 300
 3013 CAACGAGCTTCGACCTGACAAACAGTACAAACCTGAAGAGAGCTTCGACAG 3072
 301 CAGTATATCTTTGATTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 360
 3073 AAGAGACAGCTGCGCAAGTATAGATGAGATCAAGAGTGAAGTCTGCTCAAGAGCAG 3132
 361 AAGAAACAGCTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
 3133 CTTGAAGAACTCACTCACTGAAACATCGGAACAGTACGAGAACTTCGAGG 3192
 421 TTAGATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
 3193 TTCTTCAACAGAGAGAGAGCCGAGATGCGGAGACAGAAACACTGTGAGAGACAC 3252
 481 TTCTTCAACAG 540
 3253 AAGATTTCTTCAACAG 3312
 541 AAAAAATATTTGAACATTAATTAAGAGAGCTTTAAATTAATTAATTAATTAATTAATTA 600
 3313 CTGAAGAGCTCTCTCGAGAGAGAGATCCAGACCGAGAGATTAAGTACGCCAGCTCGAAG 3372

601 TTAATAACTTTAAGTGAAGAAATCAATTCAAACAGAGATTAATTAATTAATTAATTA 660
 3373 TTCAGGCTCTCTGTAAGCTGCAAGGACAGCGAAGGACAACTGAGCTGAGAGAGAG 3432
 661 TTTAAAGTATTAAGTAAATTAAGAAATTAAGAAATTAATTAATTAATTAATTAATTA 720
 3433 AAGCTCAGCTACCTCTCTGACGAGCTGACATCTGATGCGGAGCTCAAGAGATCAT 3492
 721 AATATATCACTATTAATCAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
 3493 AAGACAGAGACTACACCGGCAATGACCAAGCAGAGATTAATTAATTAATTAATTAAT 3552
 781 AAAAAATTAATAATTAACAGATTAATTCACAGTGAAGAAATTAATTAATTAATTAAT 840
 3553 CAGGATCTTACAGAGAGTCTCGCTGAGAGACAGATGTCGCCATGTGGTCTGAA 3612
 841 TTAGATCTTACAGAGAGATTTCTCCAGAGAGACAGATGTCGACAGTGTAACTGAA 900
 3613 TCTGCTCCGACACACTGAGAGAGTCTCAACCTAAGAGAGCTGATCTACATGTCGA 3672
 901 AGTGAATCCGACACATTAAGCAAGTCAACCAAGAAACAGATCACTCATGATGAGA 960
 3673 GCGAGTCCATATGACATTAACATCTGACAGACCTGAGAGATGAGTGCATGACGTCAC 3732
 961 GGAGAGCTTACACCAATTAACATCAACAAATGTCGATGATGAATGATGATGATGAT 1020
 3733 ATTGTCCTATCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3791
 1021 ATAGTATCTATTAATTTGAGAAATCCAGAGAGATTAATTAATTAATTAATTAATTA 1079

RESULT 39
 PFMEZSAID
 LOCUS P.falciparum (RO-71) gene for merozoite surface antigen 1.
 DEFINITION Z35329.1 GI:535257
 VERSION Z35329.1 GI:535257
 KEYWORDS merozoite surface antigen 1 gene.
 SOURCE Plasmodium falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 1956)
 Tolle,R., Bujard,H. and Cooper,J.A.
 Plasmodium falciparum: Recombination within the C-terminal region
 of merozoite surface antigen-1
 Unpublished
 2 (bases 1 to 1956)
 Olafsson,P., Matile,H. and Certa,U.
 Plasmodium falciparum: the repetitive MSA-1 surface protein of the
 RO-71 isolate is recognized by mouse antibody against the
 nonrepetitive repeat block of RO-33
 Exp. Parasitol. 74 (4), 381-389 (1992)
 92275047
 1592091
 3 (bases 1 to 1956)
 Tolle,R.
 Direct Submission
 Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum
 fur Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer
 Feld 282, Heidelberg, 69120, Germany
 4 (bases 1 to 1956)
 Tolle,R., Bujard,H. and Cooper,J.A.
 Plasmodium falciparum: variations within the C-terminal region of
 merozoite surface antigen-1
 Exp. Parasitol. 81 (1), 47-54 (1995)
 95354793
 762856
 FEATURES
 source
 Location/Qualifiers
 1..1956
 /organism="Plasmodium falciparum"
 /isolate="RO-71"
 /db_xref="taxon:5833"


```

/gene="gp190, MSA1, MSP1, PMMSA"
<1.1956
/gene="gp190, MSA1, MSP1, PMMSA"
CDS
/db_xref="GI:535258"
/product="merozoite surface antigen 1 gene, 3+-end"
/db_xref="GI:535258"
/db_xref="SPTREMBL:O25924"
/translacion="FNKRELGDQKMOIKRLTLLEQLESKLSLNPVYLVNLFVSF
FNKREKLEIETENTLENTLILKHYGLVLYNGESSPLKLESEVSIQEDNANKE
KPRALSKIDKLNLDNHLKRLKLSFLSGHLHLYTELKVNKNNTYSPENNKVY
NEAKSYENFLPEAKVTVTVPPOPDPVTPSPSLVRYSGSSGSTEETQIPSSGLFE
LOOVOLONYDEEDDSLVLPPIGSESDNDEYLDVYVTEGASISVTMDNISGPENED
VYLKPLAFYRSKROIEKNITFNINLNDILNSRLKKRYFLDYESIDLMOKHIS
SNEYILLESFKLNSKQKNTLSYKIKRESVENDIKFADEGISYERVLAKYKDDLE
SIRKVIKEEKEFPSSPPTPPSPAKTDEOKESKFLPELNIETLVNINLVKIDYLI
NLKAKINDCNEVERDEAHVKITKLSDLKAIKIDLEFNHNDFAEIKLINDTQKDL
GLISTGLVONFNTIISKLEGFODMLNISOHCYKQCCPOMSGCFRHLDEBECK
CLINTYQEDKCYENPMPNENNGGCDNAKCTEEDSGSNGKRTTECTKRPDSYPLF
DGIFCSSNLFGLSFLILMLILYSFT"
BASE COUNT      855 a      263 c      265 g      573 t
ORIGIN
Query Match      11.1%; Score 546.4; DB 3; Length 1956;
Best Local Similarity 57.8%; Pred. No. 3.5e-109;
Matches 1134; Conservative 0; Mismatches 726; Indels 102; Gaps 5;
QY 3064 TTCGACAGAGAAGAGACAGTCGCGAGATATAGATCGATCAAGAGTGTACTCTGCTC 3123
Db 1 TTTATATAGAAAAAAGAACTTGCCAGACAAATGCAATTAATAAACTTACTTTATTA 60
QY 3124 AAGAGACAGCTTGAACAACTCACTCACTGAACAATCCGAAACAGCTACTGACAGAC 3183
Db 61 AAAAAACAATTAAGATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120
QY 3184 TTTCAGTGTCTTCAACAAGAGAAGAGACCGGATCGCGAGACAGACACACTCTG 3243
Db 121 TTTTCTGTTTCTTTAAACAAAAAAGAAAGAACTGAATAGAGAACTGAAACACATTA 180
QY 3244 GAGAAGACCAAGATCTCTCAACACTACAAAGCCCTCGTCAATATATATATATATATAT 3303
Db 181 GAAAAACAAAAATATATTATGAACATTATTAAGACTTGTAAATATATATATATATATAT 240
QY 3304 TCTTCTCTCTGAGAGACTCTCTCGAGAGACGATCCAGACCGAGATACTACGCCAGC 3363
Db 241 TCATCTCCATTAAAAAACTTTAAGTGAAGTATCAATTCACAAAGAAATATATATG 300
QY 3364 CTCGAGAACTTCAAGTCCTGCTAAGCTGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3423
Db 301 TTTAGAAAAATTTAGAGCATTTAGTAAATAGATGAAAACTCAATGATTAATTTACATTTA 360
QY 3424 GAGAAGAGAAGAGCTACACTACTCTTACGCGAGCTGCATCTGATCGCGAGACTCAAG 3483
Db 361 GGAAGAAAAAATATATCTTTCTTATCAAGTGATTTACATCTTTAATTTAGCAATTA 420
QY 3484 GAAGTCATTAGAACAAGAACTACACCGGCATATAGCCCAAGCGAATATATACAGAGCTG 3543
Db 421 GAATTAATTAATAAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480
QY 3544 AATAAGCACTGGAATCTTACAGAAGTCTGCTGAGGAAGCAAGATGTGGC----- 3596
Db 481 AAGCAAGCTTTAAATCTTACGAAAAATTTTCTCCAGAGCAAGAAAGTTACAGAGTTGTA 540
QY 3597 -----CACTGTGTCTGGAATCTGCTCCGACACACATGGA 3632
Db 541 ACTCACCTCAACAGATGTATCTCATCTTCATTATCTGTAAGGTATGTAAGTGTCA 600
QY 3633 GCAATCTCAACCTTAAGAA-----GCCTGATCTACTCATGTGCGAGCCGAGTCCAT 3684
Db 601 GGATCCACAAAGAAAGAACAAATATCAACTTCAGGCTCTTTATTAACAGAAATTTACAA 660

```

```

QY 3685 ACAATTACCACTCAGACAGTGCAGATGAGTGCATGATGATGATGATGATGATGATGATG 3744
Db 661 CAGTAGTACATTTACAAAATTTATGACCAAGAGATGATGATGATGATGATGATGATGATG 720
QY 3745 TTGCGCGAGACGAGAGAGACTACGATGACCTCGCCAGGTGTCACCGGTGAGGCTGTC 3804
Db 721 TTTGAGAAATCCGAAGATATAGCAATATTTAGATCAAGTAGATGATGATGATGATGATG 780
QY 3805 ACTCTTCCTGATTTGATTAATCTGTGCAAAATTCGAAGCAATACGAAGTGTCTAT 3864
Db 781 TCTGCACAATG---GATTAATATCTCTCAGAGATTTGAAAAATGAAATGATGATTAATAT 837
QY 3865 CTGAACCTCGCGAGGCTGTATAGTCTCTCAAGAAACGTCGCAATACGATGATGATG 3924
Db 838 TTTAAACCTTTAGCTGGAGTATATGAAAGCTTTAAAAAACAAATTTGAAAAAACCTTTT 897
QY 3925 ACCTTCAATGTCAACGTAAGGACATTTCTGAACAGCCGCTTTAATAGAGAAAAATTTG 3984
Db 898 ACATTTAATTTAATTTGACGATATCTTAATTCACGCTTTAAGAAACGAAAAATTTTC 957
QY 3985 AAGAAAGCTTGTGAGAGAGCTGATTCCTATTAAGCTGAGCTCTCTAAGCTGT 4044
Db 958 TTAGATGTATTAGATCTGATTTATGCAATTTAAACATATATCTCTCAAAATGAAATCA 1017
QY 4045 GTCAAGACCCATCAACAGTCTCTCAATTAAGGAAGAGAGATTAATTTCTGCTAGTATAC 4104
Db 1018 ATTGAGATTTATTTAATTTATTTGATTCAGACAAAAAACACACTTTTAAAAAGTTAC 1077
QY 4105 AACTATATCAAGAGACTCCATGACACCGATATCAATTTTCGTAATGATGTCTGGGAT 4164
Db 1078 AATATATATAAAGATCGTGAATAATGATATTAATTTGACAGAGAGATTAATGAT 1137
QY 4165 TACAAGATCTCGAGGAAAAATACAGTCTGACCTTGTCTATTTAAAAAGTATATCAAC 4224
Db 1138 TATGAAAGGTTTATGCGAAATATAGATGATTTAGATCAATTAATAAAGTTATCAAA 1197
QY 4225 GATAAG----- 4230
Db 1198 GAAGAAAAAGAGTTCCTCATCTACACCAACCAACACCTCCGTCAACGCAAAAAAGAC 1257
QY 4231 ---CAAGCGAAGATGAATAATATCTGCTCTGTAATATCAATGAAACCTGTACAG 4287
Db 1258 GAACAAAAAGAAAGATGATGATCTCTCTCTTAAACAAATGAGACCTTATACAT 1317
QY 4288 ACAGTGAACGCAAAATGACCTCTGTAATATCACTGGAGGCGCAAGTCTCTACATAT 4347
Db 1318 AACTTAGTTAATTAATAATTTGACGATTTACTTAATTAACCTTAAAGGCAAAAGATTAC 1377
QY 4348 ACTTAGGGAAGACCAATGTGGAAGTTAAATCAAGGAGCTGAACCTACCTCAAAACATC 4407
Db 1378 AATGTTGAAAAAGATGAAGCACTGTTAAATTAATTAATTAATTAATTAATTAATTA 1437
QY 4408 CAAGACAGCTGCGAGATTTTCAAGAAAAATTAACATTTTCTGCGAATTTGACAGCTGTCT 4467
Db 1438 GATGACAAATATAGATCTTTTAAAAACCATATACGCTTCGAAGCAATTAATAAATTTGATA 1497
QY 4468 ACCGATTTAATCAACCAACATCTCTGACCAAGTTTCTGTCACTGCGATGTGTTCGAA 4527
Db 1498 AATGATGATGAGAAAAAAGATATGCTTGGCAAAATTAATTAATTAATTAATTAATTA 1554
QY 4528 AACCTGCGCAAAAGCTGCTGACCAATCTGCTGACGCAACCTGACGAGGAGCTGTCAAC 4587
Db 1555 AATTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1614
QY 4588 ATCTCCAGCAACATGCGTGAAGAAACAGTGCACCAAGATTAAGCGGTGTTTCAGGAT 4647
Db 1615 ATTTCACAAACCAATGCGTAAAAAACAAATGTCACAAATTTCTGATGTTTCAGACAT 1674
QY 4648 CTGACGACGCGCAAGAGTGAATGTCTCTGAACATCAACAAAGAAAGATTAAGTGC 4707
Db 1675 TTAGATGAAGAAAGAAAGATGTAATGTTTATTAATTAACAAACAAAGAGGATTAATATGT 1734

```

QY	4708	GTGGGAACCCAAACCCCTACCTGCAATGGAAACCAATGGGGGGGTGGAGCCGACTCTAA	4767
DB	1735	GTGGAAATCCAAATCCCTACTCTGTACGCAAAATTAATGGGTGATGACATGCCA <td>1794</td>	1794
QY	4768	TGCACCGAGGAAGACAGCGGCTCTAACGGAAGAAATACATGCGAGTGTACTAAGCC	4827
DB	1795	TGTACCGAAGAAATTCAGGTAGCAAGGAAAGAAATACATGTGAATGTCTAAACT	1854
QY	4828	GACTCTCATCCACTTGTGACGCGGATTTTTCCTCCAGCTCTAATTTCTGGGCATTC	4887
DB	1855	GATTTCTATCCACTTTTGCATGTGATTTTCGACAGTCCCTAACCTCTTACGAATATCA	1914
QY	4888	TTCCCTGATCCCTCATGCTGATCCCTGATACAGTTCATCA	4929
DB	1915	TTCTATTATATCACTCATGTATTAATATTATACAGTTTCATTTAA	1956
RESULT	40		
LOCUS	PFMSA1AH	999 bp	DNA
DEFINITION	Plasmodium falciparum clone 828 major merozoite surface antigen (p190) gene, partial cds.		
ACCESSION	M77720		
VERSION	M77720.2	GI:21040178	
KEYWORDS	T-cell epitope; major merozoite surface antigen; tripeptide repeat.		
SOURCE	Plasmodium falciparum.		
ORGANISM	Plasmodium falciparum.		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 999)		
TITLE	Jongwutiwes,S., Tanabe,K., Nakazawa,S., Yanagi,T. and Kanbara,H.		
JOURNAL	Sequence variation in the tripeptide repeats and T cell epitopes in p190 (MSA-1) of Plasmodium falciparum from field isolates		
MEDLINE	Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)		
PUBMED	92228018		
REFERENCE	1373473		
AUTHORS	2 (bases 1 to 999)		
TITLE	Jongwutiwes,S.		
JOURNAL	Direct Submission		
COMMENT	Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn University, Ramad, Pathumwan, Bangkok 10330, Thailand		
FEATURES	On May 21, 2002 this sequence version replaced gi:309707.		
source	Location/Qualifiers		
	1..999		
	/organism="Plasmodium falciparum"		
	/db_xref="taxon:5833"		
	/map="g1"		
	/clone="828"		
	/sex="asexual"		
	/dev_stage="erythrocytic"		
	<1..>999		
	/gene="p190"		
	<1..>999		
	/gene="p190"		
	/codon_start=1		
	/product="major merozoite surface antigen"		
	/protein_id="AAA29667.2"		
	/db_xref="GI:21040179"		
	/translation="ALEDAVLVTGYSLFQKEKMYLVNEGTSQAVTTSPFGSKSYTSSGGSVTGSGYTSVAVSAVASVASVAVSRPNSDSSDPAKSYADLKHVRNRYNLTITKELKLPOLPLTNHMLTLCNINHEFKYLDIYEINILKLYNPPLLAUKNLN DVCANDYCOIPFNKIRANEDLVAKLVGFRPLDNTKDVNGKMEYIKNNKTTIEINLEIESKRTIDKNKATKEEKRKLYQAOYDLFYNKOLEAHNLSVLEKRDITLIKNNIKNIKLDKINIKNPSPANGNPNTLLDNKKIKIEEKEIKETAKIKENID"		
BASE COUNT	421 a	125 c	150 g
ORIGIN			
QY	109	GCTTTGGAAGATGCGCTCTTACCGGATACACCTGTTCAGAAAGAGATGCTGCTG	168
DB	1	GCTTTGGAAGATGCGCTATTTGACAGTTATAGTATTATTTCAAAAGAAATGCTATTA	60

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS
PMFZSA1A	P.falciparum	Z35326	M35727	Z35326.1 GI:535247
PMFZSA1A	1956 bp	DNA	linear	INV 22-SEP-1995
PMFZSA1A	gene for merozoite surface antigen 1.			
PMFZSA1A				merozoite surface antigen 1 gene.
169	AATGAGGACGACGAGGCGACGGCGCTTACACACACACACCGGTTCTAAAGGGCTGTCG	228		
169		228		
61	AATGAGGACGACGAGGCGACGCTTACACACACACACCGGTTCTAAAGGGCTGTCG	120		
61		120		
229	GCTACCGGAGGCGGCGGCGGCGTCTGCGCTTGGGGGGTTCGTCGCTCCGGCGGACG	288		
229		288		
121	ACTTCAGGCGGTTCAGTTACTTCAGGTGGTTCAGTTACTTCAGTTGCTTCAGTTCTTCA	180		
121		180		
288	GTCGATCAGGTGCTCAGTGGCAAGCGGCTTCCGGGACAGTCGAGAACCATCCA	348		
288		348		
181	GTTGGCTCAGTTGCTTCAGTTGCTTCAGGTGGTTCAGTTAAATCAAGACGTACCAATCT	240		
181		240		
349	TTGCAACTCTAGGAGTTCCGACCCCAAGTCTACGCGGACCTCAGACAGAGTGGAGA	408		
349		408		
241	TGAGTAAATTCAGGATTCAGATCTTAATCTTACGCTGATTTAAACACAGAGTACGA	300		
241		300		
409	AACATCTCCTCCTATCAAGAGGAGCTGAATCCACAGTTGGTTGACCTCATATCAT	468		
409		468		
301	AATTCCTGTTAACTATCAACACACACTCAAAATCTCTCAACTCTTGATTTAATCAT	360		
301		360		
469	ATGCGACACGCTGATTAACATTCATGCTCAATATTCGATGAGGTTACAGAG	528		
469		528		
361	ATGTTAACTCTGTTGATTAATTCATGTTCAATTTTAAATGATGATATCAAGA	420		
361		420		
529	ATCAATGAACCTCTCTACAGAGTTGAATTTCTACTTCGACTTCGACGAGGCCAACTGAT	588		
529		588		
421	ATTAAATGAATTTATTAATAATTAACCTTTATTTTGAATTTTAAAGACAAATTAAT	480		
421		480		
588	GACGTTTGGCCCAATGACTATTGTCAATTCATTCATTTGAACATCAGAGCCACAG	648		
588		648		
481	GATGATGATGCTTAAAGATTAATGTCAAAATCCTTCAATCTTAAATTCGTGCAATGA	540		
481		540		
649	TTGACGATTAATGAAGTGGTCTCGGATTCGCAAGCTCTGACAACTCAAGGAC	708		
649		708		
541	TTAGACGTRCTTAAAAAAGCTTGTGTCGATTTACAAAAACCTTACACATTTTAAAGT	600		
541		600		
709	AATGTGGGAAAGTGAAGATTATTTAAAAAGATTAAGAAAGCAGCATGAGAACATTAA	768		
709		768		
601	AATGTAGGAAATGGAAGATTACATTTAAAAAATTAACCAATGAAATATTAAT	660		
601		660		
769	GAGCGATTCGAAGATCCCAAAAGACCCATAGCAAAATTAAGAAATGCAACCAAGAGGA	828		
769		828		
661	GATTAATTTGAAGAAAGTAGAAACCAATGTAAAAATTAAGAAATGCAACCAAGAGGA	720		
661		720		
829	GAAAGAGAAAGTGTACACAGCGCCAGTACGCTGTCCATCTATACAAACAGCTTGA	888		
829		888		
721	GAAAAAATAATTTTACCAACGCTCAATATGATCTTTTATTTACAAATTAACAAATTA	780		
721		780		
889	GAGCCCATACCTCATCAGCGTACTGGAGAGCGCATAGACACCCTCAGAGAAATGA	948		
889		948		
781	GAGACACATTAATTTAATTAAGCGTTTAGAAACGTAATGACATTTAAAAAATAATGA	840		
781		840		
949	AATATCAAAAGACCTCTGACAAAGTTAATGAATTAAGAAATCCCGCCAGCCCACTCT	1008		
949		1008		
841	AACATTAAAGGAATTAATTTGATTAAGATTAATGAATTAATTAATTAATTAATTAAT	900		
841		900		
1009	GGAACACCCCTTAACAGCTGCTGACAAAGACAGAAAGATGAGAGACAGAGAAAG	1068		
1009		1068		
901	GGAATTTACACCAATACTCTCTTATATAGAACAAAAAATGAGAGAACAGAAAAAGA	960		
901		960		
1069	ATCAAGAGATGCCCAAAACCTTAAGTTCAACATAGAT	1107		
1069		1107		
961	ATAAAAAGAAATTCACAAACCTTAATTAATTAATTAAT	999		
961		999		

[illegible]

FEATURES	source
LOCUS	AF061140
DEFINITION	Plasmodium falciparum isolate IfAI2 merozoite surface protein 1
ACCESSION	AF061140
VERSION	AF061140.1
KEYWORDS	GI:3821972
ORGANISM	Plasmodium falciparum.
REFERENCE	Plasmodium falciparum Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and Plüschke, G.
TITLE	Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania
JOURNAL	Acta Trop. 74 (1), 51-61 (2000)
MEDLINE	20106724
PUBMED	10643908
REFERENCE	2 (bases 1 to 1142)
AUTHORS	Jiang, G., Daubenberger, C.A., Matile, H. and Plüschke, G.
TITLE	Direct Submission
JOURNAL	Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel 4002, Switzerland
FEATURES	Location/Qualifiers
source	1..1142
	/organism="Plasmodium falciparum"
	/isolate="IfAI2"
	/db_xref="taxon:5833"
	/chromosome="9"
	/note="Wild isolate"
gene	72..>1142
	/gene="msp1"
mrna	72..>1142
	/gene="msp1"
	/product="merozoite surface protein 1"
CDS	72..>1142
	/gene="msp1"
	/note="p190; surface protein with spectrin binding activity"
	/codon_start=1
	/product="merozoite surface protein 1"
	/protein_id="AAC69739.1"
	/db_xref="GI:3821973"
	/translation="MKIIFPLCSFLFIINOCVTHESYOELVKKLEALDPAVLTGYS LFOEKAVLNEGTSGTAIVTTSTPESVASGGSVASGSGSVASGSGSVASGGSVAS GGSSNRRSTSPSDAKSTADIKLHVRANVLTITKEIKYPELFDLTNMMITLCON IHGKRYVITDEEINELLYKINFYFDLLIRAKLNDVCANCOIPFNKIRANEDLDYIK

Query Match	10.9%; Best Local Similarity 69.6%; Matches 745; Conservative	Score 538.8; DB 3; Length 11422; Pred. No. 1.5e-107; 0; Mismatches 322; Indels 3; Gaps 1;
10 ATGAAATCATTTCTCTCTCTGTCATCTTGGTTTTTATCATCAATACAGTGGCTG 69	137 c	365 t
72 ATGAAATCATTTCTCTCTCTGTCATCTTGGTTTTTATCATCAATACAGTGGCTG 131		
70 ACCCAAGATTCCTATCCAGAGAGCTGGTTAAGAACTGAGAGCTTGGAGATGCCCTT 129		
132 ACACATGAAAGTTATCAAGAACTGTCCAAAAACATGAAAGCTTGAAGATGCACTATTTG 191		
130 ACCGATATACGCTGTTCAGAGAGAGATGGTGTGATGAAGGACGAGTGGCAGC 189		
192 ACAGGTTATAGTTTATTTCCAAAAGGAAAAAGGATTAATTAATGAAGGAAACAAGTGGACA 251		
190 GCGGTTACACACGACACCGCGGTTCTAAAGGGGTGGGGCTAGCGGGGGCTCCGCGGG 249		
252 GCGTTATACACTAGTACACTGTGTTACAGTTGCTTAGTGGTTTCAGTTGCTTACAGTGGT 311		
250 TCTGTGGCCTCTGGGGGGTTCCTGCTCCCTCCGGCGGACGTCGTCATCAGTGGCTCAGTG 309		
312 TCAGTGGCTCAGTTGCTTTCAGTGGTTCAGTGGCTCAGTGGCTTTCAGTGGCTCAGTT 371		
310 GCAAGCGCGGCTTCGGGAGACAGTGTGAAGAACCAATTCATCTGACACTTACAGCATTC 369		
372 GCTTACAGTGGTTCAGTGGTTCAGTGGTTCAGTGGTTCAGTGGTTCAGTGGTTCAGTGGTTC 431		
370 GACGCAATTCCTTACGCGGACCTTCAAGACGAGAGAGAACTATCTCTCATATCAGTAC 429		
432 GATGCTAAATCTTACGCTGATTTAAACACAGAGTACGAAATTACTTGTTCATTTAA 491		
430 GAGCTGAATACCCACAGTGTTCGACCTCATAATCATATGCTGACACTGTGTATAC 489		
492 GAACCAATATTCGCGAAGCTCTTGTATTAACCAATCATATGTTAACTTGTGTATAT 551		
490 ATTCAATGCTTAAATATCTGTATTCAGGTTACGAGAGATCATGAACCTCTGTACAG 549		
552 ATTCAATGCTTAAATATTAATGATGATATGAAGAAATTAATTAATTAATTAATTAATTAATTA 611		
550 TTGAATTTCTATGAGACTGCTAAGGGGCAACGAACTGATGCTGGCCATATCATAT 609		
612 TTAACTTTATTTATGATTTATTAAGAGCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 671		
610 TGTCAAAATTCATTAATTTGAAGATCAGAGCCAAAGATGTCATATGTAAGAAGTTG 669		
672 TGTCAAAATTCATTAATTTGAAGATCAGAGCCAAAGATGTCATATGTAAGAAGTTG 731		
670 GTCTTGTGATTCGACAGCTCTGCACACATCAAGACATGTGGGAAAGATGGAAGAT 729		
732 GTGTTGCGATATAGAAAAACATTAACAAATTTAAAGATTAATGTAGGAAAAATGGAAGAT 791		
730 TATATTAATAAATAAAGAACGACATCGAGAACATTAACGAGCTGATGGAAGAAATCCAAA 789		
792 TACATTTAAAAAATAAATAAACAACATACCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 851		
790 AAGACCATATGACAAATAATACAAATCAACCAAGAGAGAGAAAAAGAAAGATTTGACAG 849		
852 AAACCAATTTGATCAAAATTAAGATGACATATTAAGAGAGAGAAAAAATAATTAACCA 911		
850 GCCAGATACAGCTGTCCATCTATTAACAACAGCTTGAAGAGGCCATTAACCTCATCAGC 909		
912 GCTCATATATGATCTTTATTTATTAACAATTAACAATTAACAAGAGCATAATTTAATAAGC 971		
910 GTACTGTGAGAGAGCATATGACACCTCAAGAGAAATGAATAATTAACAAGAACTGCTGCAC 969		
972 GTTTTATGAAAAAGATTTGACCTTTAAAAAATAAAGAAATTAATTAATTAATTAATTAATTAATTAATTA 1031		

Db 1018 ATGGAAGATTCATTAAATATATGCAACACAAAAACACACTTTAAAAAGTTAC 1077
 QY 4105 AACTATATCAAGACCTCATGACACGATATCATTTCCGTAAGATGTGGGGAT 4164
 Db 1078 AATATATATAAAGATCAGTGAATATATTAATTTGACAGGAAGTAAATGAT 1137
 QY 4165 TACAGATCTGAGGAAAAATACAGTCTGACCTGATATATTAAGATATCAAC 4224
 Db 1138 TATGAAAAGTTTATGCAAAATATAGATGATTTAGATCATTAATAAAAGTTATCAA 1197
 QY 4225 GATAG----- 4230
 Db 1198 GAGAAAAAGAAAGTTCCATCATCACACCAACACCTCCGTCACGCAAAAAA 1257
 QY 4231 -----CAAGGAGAAATATCTGCCCTTCTGAATTAACATGAAACCGCTGAC 4284
 Db 1258 GACGACAAAAAGAAAGTAAGTCTTCCATTTTAAACAACTGAGACCTTATAC 1317
 QY 4285 AAGACATGAACGACAAAAATGACCTCTGTAATTCACCTGAGCCAGGCTCTCAAC 4344
 Db 1318 AATACTTAATTAATAAATGACGATTAATTAATTAAGCAAAAGATTAACGAT 1377
 QY 4345 TATACTTCAGGAAGACAGATGGAATTAATAACAAGAGCTCACTACTCAAAACA 4404
 Db 1378 TGTATGTTGAAAAAGATGAAGACATGTTAAATTAACCTAAGTATTAAGCA 1437
 QY 4405 ATCCAGACAGCTGCGAGATTTTCAGAAAAATTAACAATTCGTCGGAATTCAGACCTG 4464
 Db 1438 ATGATGCAAAATATGATCTTTTAAAAACCTTACGACTTCGAAACATTTAAAAATTG 1497
 QY 4465 TCTACGATTAATACACACACATCTCTGACCAAGTTTCTGTCACCTGACATGCTGTC 4524
 Db 1498 ATAAATGATGATACAAAAAAGATATGCTTGCAAAATTAAGTATTAAGAGATTAAG--TT 1554
 QY 4525 GAAAACTCCGCAAAACAGCTGCTGAGCAATGCTCTGACGCGACACTGACAGGCTGCTG 4584
 Db 1555 CAAAATTTTCTTAATACAAATATCAAAATTAATTAAGAAAAATTCGAATATGTTA 1614
 QY 4585 AACATCTCCAGCACCATGCTGGAAGAAAGAGTCCCGCAAGATGAGGCTGTTTCAGG 4644
 Db 1615 AACATTTACACACCATGCTGTAATAAACAATTCGCAAAATTTCTGATGTTTCAGA 1674
 QY 4645 CATCTGACGAGCGGAGAGTGCATGCTCTGAACTCAAAACAAGAGAGATAG 4704
 Db 1675 CATTTAGATGAAGAGAGATGTAATGTTATTAATTAACAACAGAAAGTGATAA 1734
 QY 4705 TGCCTGGAAGAACCAACCTACCTGCAATGAAACATGCGGCTGAGAGCCGATGCT 4764
 Db 1735 TGTGTGAAAAATCCAAATCTACTGTGAAGAAAAATATGATGATGACAGATCC 1794
 QY 4765 AATGCAACGAGAAAGACAGGCTCTAAGGAAAAAATACATGCGAGTACTAG 4824
 Db 1795 ACATGTACCGAAAAAGATTCAGGTAGCAGCAAGAAAGAAATCATGTGAATGATCTAAA 1854
 QY 4825 CCGGACTCTCACTCACTCTTGCAGCGGATTTTTTGCCTCACTCAATTTCTCGGCGATC 4884
 Db 1855 CCGTGTCTTATCCACTTTTGGATGATTTTTCGCACTTCTCAATTTCTTACGATTA 1914
 QY 4885 TCTTCTCTGCTGATCTCATGCTGATCTCTGACAGCTTCACTTAA 4929
 Db 1915 TCATCTTATTAATACATCATGTTATTAATATACAGTTTCAATTTAA 1959

RESULT 45
 LOCUS AF061146 1116 bp DNA linear INV 09-FEB-2001
 DEFINITION Plasmodium falciparum isolate HN6.24 merozoite surface protein 1
 ACCESSION AF061146
 VERSION AF061146.1 GI:3821984
 KEYWORDS Plasmodium falciparum
 SOURCE Plasmodium falciparum
 ORGANISM

REFERENCE 1 (bases 1 to 1116)
 AUTHORS Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and Pluschke, G.
 TITLE Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania
 JOURNAL Acta Trop. 74 (1), 51-61 (2000)
 MEDLINE 20106724
 PUBMED 10643908
 REFERENCE 2 (bases 1 to 1116)
 AUTHORS Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
 TITLE Direct Submission
 JOURNAL Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical Institute, Soclistrasse 57, Basel 4002, Switzerland
 FEATURES
 source
 /organism="Plasmodium falciparum"
 /isolate="HN6.24"
 /db_xref="taxon:5833"
 /chromosome="9"
 /note="wild isolate"
 /gene="msp1"
 /product="merozoite surface protein 1"
 /db_xref="GI:3821985"
 /protein_id="AAC69745.1"
 /product="merozoite surface protein 1"
 /codon_start=1
 /protein_id="AAC69745.1"
 /note="p190; surface protein with spectrin binding activity"
 /gene="msp1"
 /note="p190; surface protein with spectrin binding activity"
 CDS
 gene
 mRNA
 BASE COUNT 458 a 136 c 161 g 361 t
 ORIGIN
 Query Match 10.7% Score 528.2; DB 3; Length 1116;
 Best Local Similarity 69.4% Pred. No. 3.3e-105;
 Matches 741; Conservative 0; Mismatches 308; Indels 18; Gaps 1;
 QY 10 ATGAAATCATTTTCTCTCTGTCATTTCTGTTTATCATCACTCAGTGCCTG 69
 Db 67 ATGAAGATCATATTTCTTTTATGTTTCATTTCTTTTATATATATACACATGTGTA 126
 QY 70 ACCACGAATCTATCAGAGCTGTTAAGAAATGAAAGCTTTGGAAGATGCCCTCT 129
 Db 127 ACACATGAAGATTAACAAGCTTCAAAAAGCTTTGAAGATGACAGTATTTG 186
 QY 130 ACCGATACAGCTGTTCCAGAGAGAAAGATGCTGTAATGAAGGACAGTGGCAGC 189
 Db 187 ACAGGTTATAGTTTATTTCAAAAGAAAAATGATTAATGAAGGAAAGTGGAAACA 246
 QY 190 GCCGTACACGACGACCGCGCTCTAAGAGGCTGTGGCTAGCGAGCGCGGCTGG 249
 Db 247 GCTGTTACACTAGTACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
 QY 250 TCTGTGCGCTGTGGGGGTTCCGTCGCTCCGCGGACGCTGCGATCAGTGGCTCAGTG 309
 Db 307 TCAGGTGTTCCAGTGCTTCAAGT-----TGCCTTCAAGTGTGCTCACTT 348
 QY 310 GCAAGCGCGGTTCCGGGAACAGTGAAGAAACATTCATCTGACACTCTAGCGATTCC 369
 Db 349 GCTTCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 408

[illegible]

REFERENCE	TITLE	JOURNAL	REMARK	REFERENCE	TITLE	JOURNAL	COMMENT
PUBMED	91304517		3 (bases 1 to 1009)				
PUBMED	1852173		3 (bases 1 to 1009)				
REFERENCE	Submitted (16-MAR-1994)	Ranford-Cartwright, L.C.	Direct Submission				
REFERENCE	Submitted (21-MAR-1994)	Ranford-Cartwright L.C., Institute of Cell, Animal and Population Genetics, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JN, Scotland	revised by [4] MAT				
REFERENCE	4 (bases 1 to 1009)	Ranford-Cartwright, L.C.	Direct Submission				
REFERENCE	Submitted (21-MAR-1994)	Ranford-Cartwright L.C., Institute of Cell, Animal and Population Genetics, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JN, Scotland	5 (bases 1 to 1009)				
REFERENCE	Kerr, P.J., Ranford-Cartwright, L.C. and Walliker, D.	Proof of Intragenic recombination in Plasmodium falciparum					
REFERENCE	Mol. Biochem. Parasitol. 66 (2), 241-248 (1994)						
REFERENCE	95107347						
REFERENCE	7808474						
REFERENCE	On Mar 31, 1994 this sequence version replaced g1:461333.						
FEATURES	See <52963> for MSA-1 clone 3D/A.						
FEATURES	Location/Qualifiers						
FEATURES	1..1009	/organism="Plasmodium falciparum"					
FEATURES	/db_xref="taxon:5833"						
FEATURES	/chromosome="9"						
FEATURES	/clone="HB3A"						
FEATURES	1..1009	/gene="MSP1"					
FEATURES	<1..>1009	/gene="MSP1"					
FEATURES	/codon_start=1						
FEATURES	/product="merozoite surface protein 1"						
FEATURES	/protein_id="CAA37135.1"						
FEATURES	/db_xref="GI:467604"						
FEATURES	/db_xref="SPRMBL:025865"						
FEATURES	/translation="THSYQYIKYKLLFALDAVLTGYSLFQKEMVLEGGSGAVTT						
FEATURES	STPGSKSVASGSGSGSVASGVSASGSGSRRTNSDMSDSDAKSYADLRKRVNR						
FEATURES	YLTLTKELKLPQLEFDLTNHLTLTDNTHGFRYLIDYEEINELTKLNFYDLRLARL						
FEATURES	NDVCANDYQIPEMLKIRANELDLVLRGLVGRKPLIDKNGKMDYIKRNRKTE						
FEATURES	NIELLEGSKRTIDKNNKNAPEEKKKLYQAOYDLSTYKQLEBANMLISVLEKRIDTD						
FEATURES	LKKRNIKKLIDKINEIKNPSPANGSWPTLLDKNKKIEHEKEIKIETIKFNID						
FEATURES	SL"						
BASE COUNT	437 a 125 c 153 g 294 t						
ORIGIN							
Query Match	10.7%; Score 527.8; DB 3; Length 1009;						
Best Local Similarity	70.5%; Pred. No. 4e-105;						
Matches 737; Conservative	0; Mismatches 272; Indels 36; Gaps 1;						
OY	70 ACCACGATCCCTATCAGGAGCTGTTAAGAACTGGGAAGCTTTGGAGATGCCCTCTT	129					
Db	1 ACACATGAAAGTTTACAAAGAACTGTCTCAAAAAACCTAGAAAGCTTAAAGAGATGCACTATTG	60					
OY	130 ACCGATACAGCTGTTCCGAGAGAGAAAGATGCTGAATGGAAGAGAGAGAGTGGCAGC	189					
Db	61 ACAGGTTTAACTTATTATTCAAAAGAAAAAATGCTATTAAATGAAGAGAAACAAGTGAACA	120					
OY	190 GCCGTTACACACAGCACACCCGGTTCATAAAGGTCGTGTGGCTAGCGGTGCTCCGGTGGG	249					
Db	121 GCTGTACAACTAGACACTGTGTTCAAAAGGTTCAAGTTCAGTTCAGGTCAAGTGGC	180					
OY	250 TCTGTGGCTCTGGGGGTTCCGTGCGCCCTCCGGCGGACGCGTGCATCGAGTGGGCTCATG	309					
Db	181 TC-----AGTTGCTTCAGGGGCTCACTT	204					
OY	310 GCAAGCGGCGTTCGGGAGACAGTCGAGAACCAATCCATCTGACACATCAGAGATCC	369					
Db	205 GCTTCAGGTGTTCAAGTAATTCACAGACGTACAATCTCTTCAGTAATTCAGAGTATCA	264					
OY	370 GAGCGCAAGTCTTACGCCGACCTCAAGCAGGAGTGAAGAACTATCTCTCATATCAAG	429					

Db 265 GATGCTAATCTAGCGTGATTTAAACACAGACGAAATCTGTACTATCAAA 324
Qy 430 GAGCTGAAGTACCACAGTGTGTGACCTCACTATCATGTGACACTGTGTATAC 489
Db 325 GAACCAATATCCCAACTCTGTATTAACCTAATCAATGTAATCTTGTGATAT 384
Qy 490 ATTATGCTTCAATATCTATGACGTTACGAAAGATCAATCACTCTGTACAG 549
Db 385 ATTATGCTTCAATATCTATGACGTTACGAAAGATCAATCACTCTGTACAG 444
Qy 550 TTGAATTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 609
Db 445 TTAACCTTTTATTTGATTTTAAAGCAAAATTAATGATGATGATGATGAT 504
Qy 610 TGTCAAAATTCATTCATTTGAAGTCAGCCACGATGCTGCTGCTGCTGCTGCT 669
Db 505 TGTCAAAATTCATTCATTTGAAGTCAGCCACGATGCTGCTGCTGCTGCTGCT 564
Qy 670 GCTTCGATATCGCAAGCTCTGCAACATCAAGACAAATGTGGAAGATGGAAG 729
Db 565 GTGTTGCGATATGCAAAACCATAGACATATTAAGATATGTGGAAGATGGAAG 624
Qy 730 TATATTTAAAGATTAAGACCAATGAGAAATTAACGAGCTGCTGCTGCTGCT 789
Db 625 TACATTTAAAGATTTAAAGATTTAAAGATTTAAAGATTTAAAGATTTAAAG 684
Qy 790 AAGACCATAGCAAAATTAAGACCAATGAGAAATTAACGAGCTGCTGCTGCT 849
Db 685 AAGACCATAGCAAAATTAAGACCAATGAGAAATTAACGAGCTGCTGCTGCT 744
Qy 850 GCCAGTATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 909
Db 745 GCTCAATATGATCTTCTTCAATTAACATTAAGACCAATTAATTAATTAAG 804
Qy 910 GTACTGGAAGCGCATAGACACCTCAAGAAATTAACGAGCTGCTGCTGCTG 969
Db 805 GTTATGAGAAACGATATGACCTTAAAGAAATTAACGAGCTGCTGCTGCTG 864
Qy 970 AAGATTAATGAATTAAGATTCCTGCGGACCACTGTGGAGACCCCTACACG 1029
Db 865 AAGATTAATGAATTAAGATTCCTGCGGACCACTGTGGAGACCCCTACACG 924
Qy 1030 CTGCAAGACAGCAAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1089
Db 925 CTGATTAAGACCAAAATATGAGAGACAGCAAAAGAGAGAGAGAGAGAGAG 984
Qy 1090 ATTAAGTTCAACATAGATTCCTCT 1114
Db 985 ATTAATTTATTTATGATGATTTAT 1009

RESULT 47
AF061136 1150 bp DNA linear INV 09-FEB-2001
LOCUS AF061136
DEFINITION Plasmodium falciparum isolate IFa9.2 merozoite surface protein 1
(mspi) gene, partial cds.
ACCESSION AF061136
VERSION AF061136.1 GI:3821964
KEYWORDS Plasmodium falciparum.
SOURCE Plasmodium falciparum.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 1150)
AUTHORS Jiang, G., Daubenberger, C.A., Matile, H., Tanner, M. and Pluschke, G.
TITLE Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania
JOURNAL Acta Trop. 74 (1), 51-61 (2000)
MEDLINE 20106724
PUBMED 10643908
REFERENCE 2 (bases 1 to 1150)

AUTHORS Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
TITLE Direct submission
JOURNAL Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel 4002, Switzerland
FEATURES
source
1. .1150
/organism="Plasmodium falciparum"
/isolate="IFa9.2"
/db_xref="taxon:5833"
/chromosome="9"
/note="wild isolate"
gene
71. .>1150
/gene="mspi1"
71. .>1150
mrna
71. .>1150
/product="merozoite surface protein 1"
71. .>1150
CDS
71. .>1150
/note="p190; surface protein with spectrin binding activity"
/product="merozoite surface protein 1"
/protein_id="AAC69735.1"
/db_xref="GI:3821965"
/translation="MKIIFLCFLFFIINTOCYTHESYOELVKLEALDEAVLTGGLFHEKMWLNCGTGTAVTSTPSKSGSVSGSGSVASGSGSVASGSGSVASGSGSVASGSGSNRRTNSDSDASVDKSKYDRIKRYONVPTIKELVPELFDLNLTLSDNHGFKYLIDGEBEINELIKNFEDLLRAKLVNVCADYQIIPFNKIRANEIDVLKLVFGYRPLDNIKDNVGMEDYIKRNTTIANINELIESKTIIDOKNADNEGKKRLYOAYDPLSTYLNKOLEAHNLISVLEKRIDTLKRNKIKLEIDITKTDAERPTTESKPNMLLDKNKRIEHEKEIKEI"
BASE COUNT 467 a 139 c 176 g 368 t
ORIGIN

Query Match 10.7%; Score 527.2; DB 3; Length 1150;
Best local similarity 68.8%; Pred. No. 5.5e-105;
Matches 743; Conservative 0; Mismatches 328; Indels 9; Gaps 1;

Qy 10 ATGAATAATCATTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 69
Db 71 ATGAAGATCATATCTTTTATGATTCATTTCTTTTATTAATAATACAAATGTGTA 130
Qy 70 ACCCAGATTCCTATCAGAGCTGTTAAGAAATGAAAGCTTTGGAAGATCCGCTCT 129
Db 131 ACACATGAAGATTAACAGAGCTGTCAAAAACAGAGAGCTTTGAAGATGACATATTG 190
Qy 130 ACCGATACAGCTGTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189
Db 191 ACAGGTATGTTTATTCATAGGAAAAAATGATTAATGAAGAAACAGTGAACA 250
Qy 190 GCCATTACACAGACACCCGGTCTTAAGAGCTGTGCGTACGAGGCTCCGCTGG 249
Db 251 GCTGTTCAACATAGTACACCTGGTTCAAAGGGTCTAGTTACTCAGGAGGTTCAAGTGT 310
Qy 250 TCTGTGGCCTCTGGGGG-----TCCGTGCGCTCCGGCGAGCGAGCTGCACAGT 300
Db 311 TCAGTTCCTTCAAGGTGTTCAAGGTGCTCAAGTGTGTTAAGGTGCTCAGTTCTCAGGT 370
Qy 301 GGCTCAGTGGCAAGCGGGGTTCCGGGAACAGTGAAGAACCAATCATCTGACAACTCT 360
Db 371 GGCTCAGTGGCTTCAAGGTGTTCAAGGTATTAACAGACCTTACAAATCTTCAGATTAATCA 430
Qy 361 AGCGATTCCGACGCCAAGTCTTAGCCCACTCAAGCACAGAGAGAGAGAGAGAGAG 420
Db 431 AGTGATTCAGATGCTAAATCTTACGCTGATTTAAACATTAAGATTAAGATTAATTAATTA 490
Qy 421 ACTATCAAGAGCTGAAGTACCACTGTTGCACTCAATCAATCAATCAATCAATCA 480
Db 491 ACTATTAAGAACATTAATTCAGCACTCTTTGATTTAACCATCATATGTAACTTTG 550
Qy 481 TGTGATAATTCATGCTCAATATATCTGATTTGACGCTTCAAGAGATCAATGAAGCTC 540
Db 551 AGTATATATATTCATGCTTCAATATTAATTAATTAATTAATTAATTAATTAATTAATTA 610

QY 541 CTGTACAGTGTGAATTTCTACTTGCATTGCTAAGGCCCAACTGATGACGTTTGCGCC 600
 Db 611 TTTATATAAATAAATCTTTATTTTGTATTTATAGACCAAAATTAATGATGTATGTC 670
 QY 601 AATGACTATTGTCAAAATTCATTCATTTGAGATGAGCCCAAGCATTTGAGCATTTG 660
 Db 671 AATGATATTATGTCAAAATACCTTTCAATCTTAAATCTGTCAATGATGATGACGTA 730
 QY 661 AAGAAATGTGCTTGGATTTGCAAGCCCTCTGACAACTCAAGCATGAGGGAAG 720
 Db 731 AAAAACTGTGTGGATTTAGAAAAACATTAAGCAATTTTAAGATTAATGTAGAAAA 790
 QY 721 ATGGAAGATTAATATTAAGAAAGATAGAACATGAGCAATTAAGCATGATGAGAA 780
 Db 791 ATGGAAGATTAATTAAGAAATTAAGAACATGAGCAATTAATTAATGATTAATGAA 850
 QY 781 GAATCCAAAAAGCAATAGCAAAATATAGATGCAACCAAGAGAGAAAGAAAGAG 840
 Db 851 GGAAGTAGAAAAAATGATGATCAAAATAGAAATGAGATATGAGAGAGAAAAA 910
 QY 841 TTGTACAGGCCAGTAGACCTGTCCATCTATTAACAAAGCTTGAAGAGCCATAC 900
 Db 911 TTATACCAAGCTCAATATGATCTTTCTATTTAATTAACATTTAGAGAGACATAT 970
 QY 901 CTCATCAGCTACTGAGAGAGCCATAGACCCCTCAAGAGAAATGAAATATCAAGAA 960
 Db 971 TTAATTAAGCTTTAGAAAAACGATTTGACATTTAAAAAAGAAACATTAAGAAAT 1030
 QY 961 CTCCTGACAGATTAATGAATTAAGAAATCCCGCCAGCCCACTCTGGAGACCCCT 1020
 Db 1031 CTGGAAGATTAATGAATTAAGAAATCAAGATGCGAAAAACCCACTGTAAGTAAAC 1090
 QY 1021 AACACCTGCTGAGAAAGACAGATAGAGAGAGAGAGAGAGATCAAGAGATC 1080
 Db 1091 AATACTCTCTTATTAAGAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGATC 1150

RESULT 48
 AF061139 1158 bp DNA linear INV 09-FEB-2001
 LOCUS Plasmodium falciparum isolate IFall1 merozoite surface protein 1
 DEFINITION (msp1) gene, partial cds.
 ACCESSION AF061139
 VERSION AF061139.1 GI:3821970
 KEYWORDS Plasmodium falciparum.
 SOURCE Plasmodium falciparum.
 ORGANISM Plasmodium falciparum.
 REFERENCE 1 (bases 1 to 1158)
 AUTHORS Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and Pluschke, G.
 TITLE Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania
 JOURNAL Acta Trop. 74 (1), 51-61 (2000)
 MEDLINE 20106724
 PUBMED 10643908
 REFERENCE 2 (bases 1 to 1158)
 AUTHORS Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
 TITLE Direct Submission
 JOURNAL Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel 4002, Switzerland
 FEATURES
 source 1..1158
 /organism="Plasmodium falciparum"
 /isolate="IFall1"
 /db_xref="taxon:5833"
 /chromosome="9"
 /note="wild isolate"
 gene 70..>1158
 /gene="msp1"
 mRNA 70..>1158

CDS
 /gene="msp1"
 /product="merozoite surface protein 1"
 70..>1158
 /gene="msp1"
 /note="p190; surface protein with spectrin binding activity."
 /codon_start=1
 /product="merozoite surface protein 1"
 /protein_id="AAC69738.1"
 /db_xref="GI:3821971"
 /translation="MKILIFLCSPLFIINTPOCVHESYQELVRLALEDAVLNGYS
 LPOKRWVINEGSTGAVTSTPSSGSGSVASGSGSVASGSGSVASGSGSVAS
 GGSVASGSGSNRTPNSDSSDASKVADLHRVQNTLFTIKELKYPELDNLNH
 LFLCDNIHFKYLIDGVEEINELLYKLYNLYDLRLAKLINDACANSQYIPNLIKIRAN
 ELDLKRIYGYRKLNDIKNDVKNMEDYIKRKTITAINIELIGSKRTIDONKNAD
 NEGRKLIYQAOYDLFTYKLOEPAHNIISVLEKRIIDLTKNENIKLEIDIDIKITD
 AKPTTESKPTLLDKNKIEHEHEIKIK"

BASE COUNT 471 a 142 c 178 g 367 t
 ORIGIN
 Query Match 10.4%; Score 512; DB 3; Length 1158;
 Best Local Similarity 69.1%; Pred. No. 1.2e-101;
 Matches 752; Conservative 0; Mismatches 315; Indels 21; Gaps 3;

QY 10 ATGAATAATCATTTTCTTCCTGCTCATTTCTGTTTATATCATCAATCACTGCGCTG 69
 Db 70 ATGAAGATCATATTTCTTTATGTTTATTTCTTTATTTATTAATATCAATATGTA 129
 QY 70 ACCCAGAAATCTATCAAGAGCTGTGTTAAGAAATGAACTTGAAGATGCGCTCTT 129
 Db 130 ACACATGAAGATTAATCAAGAACTTGTCAAAAAAATCTAGAACTTGAAGATGCAATATTG 189
 QY 130 ACCGATACAGCCCTGTTCGAGAAAGAGATGCTGTAATGAAGAGAGAGAGAGAGAG 189
 Db 190 ACAGATTATGTTTATTTCAAAAGAAAAATGTAATTAATGAAGAACAGATGAGACA 249
 QY 190 GCCGTACAAACGACACACCGGTTCTAA-----AGGCTGTGGCTACGGTGGC 240
 Db 250 GCTGTTAACAACATGATACACGCTGTTCMAAGGTTCAAGTGGCTCAGTGTTCAGTGGCT 309
 QY 241 TCCGGTGGGCTGTGGCCGTGTGGGGT-----TCCGTCCCTCCCGCGCACGCTG 291
 Db 310 TCAGGTGTTCAATTTGCTTTAGGTGGTTCAGGTGGCTCAGTGTTCAGTGGCTCAGT 369
 QY 292 GCATCAGGTGCTCAGTGGAGAGCGGCTTCCGGGAACAGTGAAGAACCAATTCATCT 351
 Db 370 GCTTCAGGTGCTCAGTGGTTCAGGTGGTTCAGGTGAATCAAGAGCTCAATTCCTTCA 429
 QY 352 GACAACCTTACGATTCACAGCCCAAGTCTACGCGGACCTCAAGACCAAGAGAGAAAC 411
 Db 430 GATTAATTCAGATGATTCAGATGTTAATCTTACGCGATTTAAACATAGAGTTCAAAAT 489
 QY 412 TATCTCTCTCATTAACAGAGGTGAAGTACCAAGTGTTCACCTCACTATTCATATG 471
 Db 490 TACTTGTTCATTTAAGAACTCAATATCCGAACCTTTTATTTAACCAATCATATG 549
 QY 472 CTGACACTGTGATTAACATTCATGGCTTCAAAATCTGATGACGGTTACGAAGATC 531
 Db 550 TTAACATTGTGTGATATATATCATGTTTCAAAATTTATTAATGATGATGAAGAAAT 609
 QY 532 AATGAACCTCTGACAGATTTGATTTCTACTTGTGACTGTAAGGCGCAACATGATGAC 591
 Db 610 AATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGAT 669
 QY 592 GTTGGGCAATGACATATGTCAAATTCATCAATTTGAAGTACGAGCAACGATG 651
 Db 670 GCATGTGCAATATGTTATTTCAAAATCTTAACTTTAAATTTGTGCAAAATGATTA 729
 QY 652 GACGATTTGAAGAGATTTGCTTCGATATCGAAGCTCTCGCAACATCAAGAGCAAT 711
 Db 730 GACGTACTTAAAAAATTTGTTCCGATATAGAAACCATTAACATATTAAGATTAAT 789
 QY 712 GTGGAAAGATGGAAGATTTATTAATAAGAAATGAAGACATCGAAGCAATTAACGAG 771

Db 790 GTAGAAAAATGGAAGATTACATTAAAAAATAAACAACCATAGCAAAATTAATATGA 849
Qy 772 CTGATCGAAGATCCAAAAAGCCATAGCAAAAATTAAGATGCAACCAAGAGAGAGA 831
Db 850 TTAATTTGAAGGAAGTAAGAAACAATTTGATGATTAAGAAATGAGATGAAGAGA 909
Qy 832 AAGAAGATGTGTACCAAGCCAGTAGACCGTCCATCTATAACAAACAGCTTAAGA 891
Db 910 AAAAAAATTTATACCAAGCTCAATATGATCTTTTATTATTAATTAACAAATTAACAGAA 969
Qy 892 GCCCATACCTCATCAGCGTACTGGAGAGCGCATAGACACCCCTCAGAAAGTAAAT 951
Db 970 GCACATATTTAATATAGAGGTTTAAAGAAAGCTATGACACTTAAAAAATGAAAA 1029
Qy 952 ATCAAGAAGCTGCTGCACAGATTATGAATTAAGA---ATCGCCGCCAGCAACTT 1008
Db 1030 ATAAAGAATTTACTGAGATATGATTAATTAACAGATGCCGAAAAACCCACTACT 1089
Qy 1009 GGAACACCCCTAACACGCTGCTGACAAAGAACAGATAGAGAGACAGAGAAAG 1068
Db 1090 GAAATTAACCAAAATATCTCTGTATAGAACAAAAAATCGAGAACACGAGAAAAA 1149
Qy 1069 ATCAAGA 1076
Db 1150 ATAAAGA 1157

RESULT 49
AF061147 1102 bp DNA linear INV 09-FEB-2001
LOCUS AF061147
DEFINITION Plasmodium falciparum isolate HN6.27 merozoite surface protein 1
ACCESSION AF061147
VERSION AF061147.1 GI:3821986
KEYWORDS Plasmodium falciparum.
SOURCE Plasmodium falciparum.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1102)
AUTHORS Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and Pluschke, G.
TITLE Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania
JOURNAL Acta Trop. 74 (1), 51-61 (2000)
MEDLINE 20106724
PUBMED 10643908
REFERENCE 2 (bases 1 to 1102)
AUTHORS Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
TITLE Direct Submission
JOURNAL Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel 4002, Switzerland
FEATURES
source
1. 1102
/organism="Plasmodium falciparum"
/isolate="HN6.27"
/db_xref="taxon:5833"
/chromosome="9"
/note="Wild isolate"
71. >1102
/gene="mspi"
71. >1102
/gene="mspi"
/product="merozoite surface protein 1"
71. >1102
/gene="mspi"
/product="merozoite surface protein 1"
71. >1102
/note="mspi"
/codon_start=1
/product="merozoite surface protein 1"
/protein_id="AAC69746.1"
/db_xref="GI:3821987"

/translation="MKIIFLCSLEFLIINTOCYTHESYDELVKLEALDEDAVLGTGS
LFOEKWLNIGTSTGATVATSTPSGSGSVTSVSGGSAVASAASVSGSGSRPTNS
DNSSDSBAKSYADLKHRYONVLPFTKELKYEFLDPLTNHMLTIDCNHGFYLDGYE
EINELDKLNFYDILRAKLNDCANSTCOIPENIKIRANELDVKTKVFRPLDN
IKDYNKMKEDYIKNKTTIANINELISESEPTIDONKNADBEKKIKYOYNLFY
NKLOEAMNLISVLEKRIDALKKNENIKKILKINELKNPPANSGMTPNLLDNKK
IEEHEKIKE"
BASE COUNT 455 a 134 c 158 g 355 t
ORIGIN
Query Match 10.3%; Score 508.2; DB 3; Length 1102;
Best Local Similarity 68.7%; Pred. No. 8.2e-101;
Matches 733; Conservative 0; Mismatches 298; Indels 36; Gaps 1;
Qy 10 ATGAATATCATTTCTTCTCTCTGTTTCATTTCTGTTTTATCATCAATCTAGTCGCTG 69
Db 71 ATGAAGATCATATCTCTTTTATGTTTCATTTCTTTTATATTAATACAAAGTGTA 130
Qy 70 ACCCAGAACTCTATCAGAGCGTGGTTAAGAACTGGAAGCTTTGGAAGATCCGCTCT 129
Db 131 ACACATGAAAGTTATCAGAACTGTGCAGAACTGTAAGAGCTTTAGAAATGCAAGTATTG 190
Qy 130 ACCGATACAGCGCTGTCAGAGAGAAAGATGCTGTAATGAAGAGGAGAGTGGACAG 189
Db 191 ACAGTTATAGTTATTTTCAAAAGAAAAATGATTAATGAAGAGAAACAGTGAACA 250
Qy 190 GCGTTACACACGACACACCGCGTTCTAAAGGCTGTGCGTACGCGTGGCTCGGTGG 249
Db 251 GCTGTTACAACTAGTACACCTGTTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 310
Qy 250 TCTGTGGCCTCTGGGGGTTCCGCGCTCCGCGGACGCGTGGTGGTGGTGGTGGTGG 309
Db 311 TC-----AGTGTCTTCAAGTCTTCACTT 334
Qy 310 GCACGCGCGGTTCCGGGACAGTCAGAGAACCAATCCATCTGACACTAGCGATTCC 369
Db 335 GCTTCAGAGTGTTCAGGTATTCAGAGCTACAAATCCCTTCAGATTAATCAAGTGA 394
Qy 370 GACGCCAGTCTTACGCGGACCTCAAGACAGAGTGAAGATATCTCTCATCTCAAG 429
Db 395 GATGCTAAATCTTACGCGATTTAAACATAGAGTGAAGTCAATCTGTTCACTATTA 454
Qy 430 GAGCTGAAGTACCCACAGTGTGTCACCTCACTATCTATCTGACACTGTGATAC 489
Db 455 GAACCTCAATATCCCGAAGCTCTTATTAACCATCATATCTTAACCTTGTGATATAT 514
Qy 490 ATTATGCTTCAATATCTGATTAAGAGTGAAGATCAATCAATCTCTGACAG 549
Db 515 ATTATGCTTCAATATCTGATTAAGAGTGAAGATCAATCAATCTCTGACAG 574
Qy 550 TTGAATTTCTACTGACTGCTAGAGGCCAATGAGTGAAGTGGGCCAATGACTAT 609
Db 575 TTAACCTTTTATGATTTATTAAGAGCCAAATTAAGATGATGATGATGATGAT 634
Qy 610 TGTCAATTCCTCAATTTTGAAGATCAGAGCAAGTGGAGTATTAAGAAAGTTG 669
Db 635 TGTCAATATCTTCAATCTTAATTTGCTGCAATGATTAAGAGTACTTAAAAAAT 694
Qy 670 GTCTTCGATATGCAAGCGCTCTGACAAATCAAGAGATGCGGAAAGATGAAGAT 729
Db 695 GTGTTGATATTAAGAAACCATTAATATTAAGATTAAGATTAAGATTAAGATTA 754
Qy 730 TATATTAAGAAATTAAGAGCGCTGCAAGATTAAGATTAAGATTAAGATTAAG 789
Db 755 TACATTAAGAAATTAAGAAACCATTAATTAAGATTAAGATTAAGATTAAGATTA 814
Qy 790 AAGACATAGACAAAAATTAAGATCAACCAAGAGAGAAAGAAAGATTTGACAG 849
Db 815 AAAAAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 874
Qy 850 GCCCAGTACGACCTGCTCATATTAACAAACAGCTTGAAGAGCCCATACCTCATCAGC 909
Db 875 GCTCAATATTAATCTTTTATTAACATTAACAAATTAAGAGCAAGATTAATTAATTAAGC 934

REFERENCE 1 (bases 1 to 1124)
AUTHORS Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and Pluschke, G.
TITLE Sequence diversity of the merozoite surface protein 1 of *Plasmodium falciparum* in clinical isolates from the Kilombero District, Tanzania
JOURNAL Acta Trop. 74 (1), 51-61 (2000)
MEDLINE 20106724
PUBMED 10643908
REFERENCE 2 (bases 1 to 1124)
AUTHORS Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
TITLE Direct Submission
JOURNAL Submitted (22-Apr-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel 4002, Switzerland
FEATURES
source 1..1124
/organism="Plasmodium falciparum"
/isolate="HN3"
/db_xref="taxon:5833"
/chromosome="9"
/note="wild isolate"
72..>1124
/gene="msp1"
72..>1124
/gene="msp1"
/product="merozoite surface protein 1"
72..>1124
/gene="msp1"
/note="p190; surface protein with spectrin binding activity"
/codon_start=1
/product="merozoite surface protein 1"
/protein_id="AAC69743.1"
/db_xref="GI:3821981"
/translation="MKIIFELCSPLFIINTOCVTHESYQELVYKKEAEADAVLTGYS
LFOKEKMLNEGSGTAVTTSTPGSKSGSGSGSVASVSGSVASVSGSGNS
RRTPNSDSDDASADKLHRYOVNLTLELKPFLDLTNHLTLCDSINHGKY
LIDYEIEINELLYKLNFYDLRAKLNDACANSYCOIPNELIRANLIDLKRIYGY
RRPLDINKDNVGMEDYIKRKTITIANINELIEGSKRTIDCKNADNEGKKLYAQ
VRLFYKNOLOEHNHLSVLEKRIDPLKKNENIKLEIDIDIKTDAEKPTGSKRNT
LIDKNKKIEHEHEKIKE"
BASE COUNT 465 a 135 c 166 g 358 t
ORIGIN
Query Match 10.1%; Score 500.8; DB 3; Length 1124;
Best local Similarity 68.4%; Pred. No. 3.5e-99;
Matches 732; Conservative 0; Mismatches 317; Indels 21; Gaps 2;
QY 10 ATGAATAATCATTTTCTCTCTGTTCAATTTTATTCATCAATCACTAGCGCTG 69
DB 72 ATGAAGATCATTTTCTTTATGTCTTCTTTTATTTATTAATACACATGTGTA 131
QY 70 ACCCAGCAATCATCAAGAGCTGTGTAAGAAGCTTGAAGATGCGGCTT 129
DB 132 ACACATGAAGATTAATCAAGAACTTTCAGAACTTTAGAGATGCAATTTG 191
QY 130 ACCGATACAGCCTGTCCAGAGAGAGAGATGGTGTGAATGAAGGACAGTGCACG 189
DB 192 ACAGGTATATGTTATTTTCAAAAGAAAAATGTTATTAATGAAGAACAAAGTGACA 251
QY 190 GCCGTATCAACACAGACCCGGTCTTAAAGGCTGTGTGGCTAGCGGTGCTCGGTGGG 249
DB 252 GCTGTATACACTAGTACACCTGGTTCAAAGGTTCAAGTGTGTCAGTGGTTCAGTGGC 311
QY 250 TCTGTGGCCTCTGGGGTTCGCTCGCTCGGGCGGACGCTGCATCAGTGTGCTCGT 309
DB 312 TCAGTGTCTTCAGTGTGCTTGAAGTGGC-----TCAGTGTCTTCAGT 353
QY 310 GCAAGCGCGGTCTCGGAGACAGTCAGAACCAATCATCTGACACTTAGCGATTCC 369
DB 354 GCTTCAGGTGTCTCAGTATTTCAAGACGTCACAAATCTCTCAGATTAATTCAGATTC 413

QY 370 GACGCCAAGTCTACGCCGCTCAAGACACCGAGTGAGAAACTATCTCTCACTATCAG 429
DB 414 GATGCAATATCTTACCTGACCTGATTTAAACATAGAGTTGACATCTGCTCACTATTA 473
QY 430 GACCTGAGTACCCACAGTGTGTGACCTGACATCATATGCTGACACTGTGTATAC 489
DB 474 GAACCTCAAAATATCCGAACTCTTGTGATTAACCAATCATATGTAACCTTGTGTATAT 533
QY 490 ATTCAATGCTTCAAAATATCGATGAGGTTTACGAGATCAATCAATCTCTGTCAAG 549
DB 534 ATTCATGCTTCAAAATATTAATTTATGATGATATGAAGAAATTAATATATATATTA 593
QY 550 TTGAATTTCTACTTCGACTGTGAAGGCCAAACTGATGACGTTTCCCAATGACTAT 609
DB 594 TTAACTTTATTTATATTTATTTAATGAAGCCAAATTAATGATCATCTGCAATATGTTAT 653
QY 610 TGTCAATTTCCATTCATTTGAAGATCAGAGCCACAGAGTGGACGATTTGAAGAAGTTG 659
DB 654 TGTCAATTTACCTTCAATCTTAATAATTCGTGCAAAATGAATTAACGCTTAATAAATTT 713
QY 670 GTCTTGATATCGCAAGCCTCTCGACACATCAAGACATGTGGGAAAGATGAGAT 729
DB 714 GTGTGGATATAGAAAGACCATTTGACAAATATTAAGATATGTAGGAAATGGAAGAT 773
QY 730 TATATTAAAAAGATTAAGAGACCATTCGAGACATTAACGACTGATCGAAGATCCAA 789
DB 774 TACATTAAAAAAATTAACACACCATAGCCAAATTAATTAATTAATTTGAAGAAAGTAA 833
QY 790 AAGACCATAGCAAAATTAAGATGCAACACAGAGAGAGAGAAAGAAAGATTTGACAG 849
DB 834 AANAACATTTGATCAAAATTAAGATGCAATGCAATTAAGAGAGAGAGAGAGAGAGAGAG 893
QY 850 GCCCAGTACAGCCTGTGTCATTAACAAACACCTTGAAGAGCCCATTAATCTATCAGC 909
DB 894 GCTCAATATATATCTTTTATTTATTAACAATTAACCAATTAACAGACATATTTAATAAGC 953
QY 910 GTCTGAGAGAGCGCATAGACACCTTCAGAGAAATGAATATCAAGAGACGTCGAC 959
DB 954 GTTTTGAAAAAAGTATTCACCTTTAAAAAAGAAAAAAGAAAAAATTAAGAAATTCAGAA 1013
QY 970 AGATTAATGAATTTAAGA---ATCCCTCGCCAGCCCACTCTGGAACACCCCTAACAG 1026
DB 1014 GATATAGATTAATAATTAACACAGATGCCGAAAAACCCACTCTGGAAGTAAACCAATACT 1073
QY 1027 CTGCTGACACAGACAGAAAGATAGAGAGACGACGAGAAAGATTCANAGA 1076
DB 1074 CTCTTGATTAAGAAACAAAAATCGAGAGACGAGAGAAAAATAAAGA 1123
RESULT 53
AF061138 1149 bp DNA linear INV 09-FEB-2001
LOCUS Plasmodium falciparum isolate IFAl0 merozoite surface protein 1
DEFINITION (msp1) gene, partial cds.
ACCESSION AF061138
VERSION AF061138.1 GI:3821968
KEYWORDS
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE 1 (bases 1 to 1149)
AUTHORS Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and Pluschke, G.
TITLE Sequence diversity of the merozoite surface protein 1 of *Plasmodium falciparum* in clinical isolates from the Kilombero District, Tanzania
JOURNAL Acta Trop. 74 (1), 51-61 (2000)
MEDLINE 20106724
PUBMED 10643908
REFERENCE 2 (bases 1 to 1149)
AUTHORS Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
TITLE Direct Submission
JOURNAL Submitted (22-Apr-1998) Molecular Immunology, Swiss Tropical

Institute, Socinstrasse 57, Basel 4002, Switzerland

FEATURES
source
1..1149
 /organism="Plasmodium falciparum"
 /isolate="IFA10"
 /db_xref="taxon:5833"
 /chromosome="9"
 /note="wild isolate"
70..>1149
 /gene="msp1"
70..>1149
 /gene="msp1"
70..>1149
 /product="merozoite surface protein 1"
70..>1149
 /gene="msp1"
 /note="p190; surface protein with spectrin binding activity"
 /codon_start=1
 /product="merozoite surface protein 1"
 /protein_id="AAC69737.1"
 /db_xref="gi:3821969"
 /translation="MKITFFLCSEFLFIINOCYTHESYQELVKLELDELDAVLTGIC
LEFKEMKLNIGSTGTAATVITSTPSGSGSVTSVSGSVASGSGSVASGSGSVAS
GGSVASGSGSRRNPDSNDSDAKSYADLKHVQVNLFTIKELKYPDELITNM
LTLCDNHGFEKYLIDGYEINELLYKLNFYDLRAKINDYACANDYCOIPNLKIRAN
ELDVLRKLVFGRPLDNIKDNVEMEDYIKKKRTTININELIGESKRTIDONKND
NEEKKRKYONQYDLSIYNKOLEAHNLSISLEKRIIDLKKNENKILLEDIDIKTID
AEKLTGSKRPLDENKKEVEGHE"
BASE COUNT 455 a 142 c 184 g 368 t
ORIGIN

Query Match 10.1%; Score 500.2; DB 3; Length 1149;
Best Local Similarity 68.1%; Pred. No. 4.7e-99;
Matches 735; Conservative 0; Mismatches 323; Indels 21; Gaps 2;

OY 10 ATGAAATCATTTCTCTCTGTCATTTCTGTTTTTATCATCAATACAGTCGCTG 69
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 70 ATGAGATCATATCTCTTTTATGTTTCATTTCTGTTTTTATTAATACACAATGTGA 129
OY 70 ACCCGAGATCCTATCAGAGCTGTTAGAACTGGAAGCTTTGGAACATCCGCTCT 129
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 130 ACACATGAAAGCTATCAAGAACTGTGCAAAAAGAGCTTTAGAAAGATGCAATTG 189
OY 130 ACCGATACAGCTGTTCCAGAGAGAGATGCTGAATGAAGAGGAGAGTGGCAG 189
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 190 ACAGTTATGTTTATTTTCATAGGAAAAATGATCTTAATGAAGAACAGTGGACA 249
OY 190 GCCGTTACACAGCAGACCCGCTTCTAAGAGGCTGTGCTAGCGGTGCTCCGCTGG 249
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 250 GCTGTTACAACTAGACACTGCTCAGGTGCTGATGTTTAAACATAGAGTCAAAAT 309
OY 250 TCTGCGCTCTGGGGGTTCCGCTCC-----TCCGGCGGACAGCTG 291
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 310 TCAGTTGCTTCAGGTGCTCAGGTGCTCAGTTGCTTCAAGTGTGCTCAGGTCAAT 369
OY 292 GCATCAGGTGCTCAGTGGCAGCGGCTCCGGAAACAGTCAAGAACCAATCATCT 351
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 370 GCTTACAGGTGCTCAGTGGCTTCAGGTGCTTCAAGCAATTCATCAATCTTCA 429
OY 370 GACAACCTAGCAATTCGACGCCAAGTCTTACGCCGACCTCAAGCAGCAGTGAANA 411
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 430 GATATTCAGGATTCAGATGCTAAATCTTACGCTGATTTAAACATAGAGTCAAAAT 489
OY 412 TATCTCCATCATATCAAGAGCTGAAGTACCAAGTGTTCGACCTCAATCATANG 471
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 490 TACTTGTTCATTAAGAACTCAAAATATCCGAATCTTGATTTAAACAATATATANG 549
OY 472 CTGACACTGTGATTAACATTCATGCTCAATATCTGATGAGCGGTAGCAAGATG 531
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 550 TTAACCTTGTGATTAATATTCATGTTCAATATTTAATGATGATATGAAGAAAT 609
OY 532 AATGACCTCTGTACAGTGAATTTCTACTGCTGCTAAGGCGCAACATGATGAC 591
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 610 AATGAATATATATTAATAAATTAACCTTTATTTGATTTATTAAGACAAATTAATCAT 669
OY 592 GTTTGGCCAAAGACATATGTCAAATTCATCATATTTAATATGAGCCAAAGCTG 651
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 670 GATATGCTTAAGATTAATGTCAAATACCTTCAATCTTAAATTCGTCAAATGATTA 729
OY 652 GACGATTTGAAGAGTGTGCTTCGATATCCGACCCCTCTGCAACATCAAGACAT 711
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 730 GACGATCTTAAAAAAGCTGTGCTGATATAGAAAAAGCTTATGACATATTTAAGATAT 789
OY 712 GTGGAAAGATGAGATTAATTAATAAAGATTAAGAACCATGAGACATTAACGAG 771
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 790 GTAGGAAAAATGAGATTAATTAATAAAGATTAATAAAGATTAATTAATGA 849
OY 772 CTGATGAGATTCACAAAAGACATAGCAAAAATTAACAAATGCAACGAAGAGAGA 831
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 850 TTAATTTGAAGAGATTAAGAAACAAATGATCAAAATTAAGATTAATGAAGAGA 909
OY 832 AAGAAGAGTGTGACAGCCAGTACGACCTGTCATATTAACAACAGCTTGAGAA 891
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 910 AAAAAAATTTATACCAAGCTCAATATGATCTTCTATTTACATTAACAATTAAGAGA 969
OY 892 GCCATTAACCTATCAGCTCTGAGAGAGCGCATAGACACCTCAAGAGATGAAT 951
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 970 GCACATTAATTTAATAAGCGTTTGAAGAAACGATTAAGCACTTTAAAAAATGAAAAC 1029
OY 952 ATCAAGAAGCTGCTGACAGATTAATGAATTAAGA---ATCCCTCCGACGCAACT 1008
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1030 ATTAAGAAATTTACTTGAAGATTAAGTAAATTAATAACAGATGCCGAAAACTCCTACT 1089
OY 1009 GGGAAACCCCTTACACGCTGCTGACAAAGACAGAGATGAGAGCAGCAAGAA 1067
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1090 GGAAGTAACCAAAATCCTCCTCGATTAAGAAAAAGAGTGCAGGACACGAGA 1148

RESULT 54
AF061137
LOCUS
DEFINITION
Plasmodium falciparum isolate IFA9.10 merozoite surface protein 1
ACCESSION
AF061137
VERSION
AF061137.1 GI:3821966
KEYWORDS
ORGANISM
Plasmodium falciparum.
REFERENCE
1 (bases 1 to 1167)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS
Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and
Pluschke, G.
TITLE
Sequence diversity of the merozoite surface protein 1 of Plasmodium
falciparum in clinical isolates from the Kilombero District,
Tanzania
JOURNAL
Acta Trop. 74 (1), 51-61 (2000)
MEDLINE
20106724
PUBMED
10643908
REFERENCE
2 (bases 1 to 1167)
Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
TITLE
Direct Submission
Submitted (22-Apr-1998) Molecular Immunology, Swiss Tropical
Institute, Socinstrasse 57, Basel 4002, Switzerland
FEATURES
source
1..1167
 /organism="Plasmodium falciparum"
 /isolate="IFA9.10"
 /db_xref="taxon:5833"
 /chromosome="9"
 /note="wild isolate"
73..>1167
 /gene="msp1"
73..>1167
 /gene="msp1"
73..>1167
 /product="merozoite surface protein 1"
73..>1167
CDS

OY	229	GCTAGCGGCGCTCGGAGGGCTCTGTGGCCCTCTGGGGGTCGTCGCGCTCCGGGCGGAC	288
Db	121	ACTTCAGGCTGGCTTCACTTACTTTCAGGTGGTTCAGTTACTTCACTTGCTTCACTTCA	180
OY	289	GTGCGCATCAGGTGGCTCAGTGGCAAGCGCGGTTCCGGGAACAGTCAAGAACCAATCCA	348
Db	181	GTTCGCTCAGTGTGCTTCAGTTGCTTCAGGTGGTTCAGGTAATTCGAAGCGTCAAAATCCT	240
OY	349	TCTGCAACCTCAGGGAATTCGAGGCGCAAGTCTCTGCGGCAACCTCAAGACCGAGTGA	408
Db	241	TCAGTAAATTCAGAGATTCAGATCTCAATCTTAATTCAGCTGATTTAAACATTAAGATTCA	300
OY	409	AACATCTCTCTCAGTATTCAGAGAGCTGAAGTACCACAGTGTGTTGACCTCAGTAATCAT	468
Db	301	AATTAAGTCTGTCAGTATTAAGAACTCAAAATATCCGAACCTTGATTTAACCAATCAT	360
OY	469	ATGCGAGACAGTGTGATATACATTTATGCGTTCAAAATTCGTATGACGGTTACGAAG	528
Db	361	ATGTTAACTTGTGTGATATTAATTCATGTTGTTCAAAATTTTAATTAAGTGAATATGA	420
OY	529	ATCAATGAACTCTGTACAGTGAAGTTGTAATTTCTACTTCGACTGCTAAAGGCCAACTGAT	588
Db	421	ATTATGAAATTAATTAATTAATTAATTAACCTTTATTTTGATTTTATTAAGACAAATTAAT	480
OY	589	GACGTTTGGCCCAATGACTATTGTCAAAATTCATTCAAATTTGAGATCAGAGCCACGAG	648
Db	481	GATGATGTGTGTAAGATTAATGTCTCAAAATACCTTTCAAATCTTAAATTCGTGCCAAATGA	540
OY	649	TTGCGATTTGCAAGAGTGTGTTGCGATTCGCAAGCTCTCGACACATCAAGAGC	708
Db	541	TTAGACGACTTAAAAAACTGTGTTCGATTAAGAAAAACCTTTGACATATTTAAAGAT	600
OY	709	AATGTGGCAAGAGTGAAGATTAATTAATAAAAAAGAAATTAAGAACCATCGAGAACATTAC	768
Db	601	AATGTGGCAAAAAATTCGAAGATTACCTTTAAAAAAAATAAACCAACCATGCAAAATTTAAT	660
OY	769	GAGCGATTCGAAGATTCAAAAAGACCTATGACAAAAATTAACAATGCAACCAAGAGGAA	828
Db	661	GAAATTAATTTGAAGAGATTAAGAAAAACAATGTATCAAAATTAAGAAATGCAATTAATGA	720
OY	829	GAAAGAGAGAGTGTACAGGCGCCAGTAGCAGCTGTCCATCTATTAACACAGCTTGAA	888
Db	721	GGAATAAAAAAATTTTACCAGACTCAATATTAATCTTTTATTTACAAATTAACAAATTACA	780
OY	889	GAAGCCCTAATCCATCAGGCTACTGGAGAAAGCGCATRAGAACCCCTCAAGAAAGATGA	948
Db	781	GAAAGCACTAATTTATTAAGCGTTTATGAAAAAAGCTATTGACACTTTAAAAAAAATGTA	840
OY	949	AATATCAAGAACTGCTGACACAGATTTAATGAATTAAGAAATCTCCGCCAGCCAACCT	1008
Db	841	AACATTAAGAAATTTCTGTATGAATTAATTAAGAAATTAAGAAATCCCCCACCGCCATCT	900
OY	1009	GGAAGACCCCTTAACAGCTGC--TGACAGAAACAAGAGATGAGAGACGACAGAAA	1065
Db	901	GGAATTAACCAAAATCTCTCCCTGAGATTAAGAAAAAAGAGTGAAGGACACAGAA	960
OY	1066	GAGATCAAGAGATGCGCCAAAAACCTTAAGTTCAACATATGAT	1107
Db	961	AAAAATTAAGAAATTTGCCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1002

RESULT	57
PfPMSALAK	
LOCUS	PfPMSALAK
DEFINITION	Plasmodium falciparum clone B35B major merozoite surface antigen (p190) gene, partial cds.
ACCESSION	M77723
VERSION	M77723.2
KEYWORDS	GI:21040184
SOURCE	T-cell epitope; major merozoite surface antigen; tripeptide repeat.
ORGANISM	Plasmodium falciparum.
REFERENCE	Plasmodium falciparum.
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE	1 (bases 1 to 1002)

AUTHORS	Jongwittiwes,S., Tanabe,K.,Nekazawa,S., Yanaqi,T. and Kanbara,H.
TITLE	Sequence variation in the tripeptide repeats and T cell epitopes in P190 (MSA-1) of Plasmodium falciparum from field isolates
JOURNAL	Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
Medline	92228018
PUBMED	1373473
REFERENCE	2 (bases 1 to 1002)
AUTHORS	Jongwittiwes,S.
TITLE	Direct Submission
JOURNAL	Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn University, Rama4, Pathumwan, Bangkok 10330, Thailand
COMMENT	On May 21, 2002 this sequence version replaced gi:309712.
FEATURES	Location/Qualifiers
SOURCE	1..1002
	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/map="g"
	/clone="835B"
	/sex="asexual"
	/dev_stage="erythrocytic"
	<1..>1002
	/gene="P190"
	<1..>1002
	/gene="P190"
	/codon_start=1
	/product="major merozoite surface antigen"
	/protein_id="AA029670.2"
	/db_xref="GI:21040185"
	/translation="ALDDAVLTGYSLFQEKWVLNEGSTGTAVTTSPPSGSVTSGGG SVMSGSVTSVAASVASVSAVSAGSGSNRRNPDSNSDPAKSYADIKRHYONY LETFKEKYLPFLDTLNMHLCINIGFKPLIDIGEINELLYKLNFYFDLLRAKLN DYNCANDCOIPENLKIRANIELDIVAKLVFGYGRRLDNIDKNVGKMEYIKNKTTIAN INEIBESKRTIDONKNADNDPEGKKRYOARYNFLYNKOLOEHNLSVLERKDIDL KNNENIKLLDKLINEIKRNPPANGNTPNLPENKKKEVBSEHEKIKEIAKTIFNID
BASE COUNT	414 a 130 c 155 g 303 t
ORIGIN	
Query Match	10.0% Score 496.4 ; DB 3 ; Length 1002;
Best Local Similarity	69.2% Pred. No.3.2e-98;
Matches 693; Conservative 0; Mismatches 306; Indels 3; Gaps 1;	
Oy	109 GCTTGGAAGAATGCCCGTCCTTACCGCATACACGCTGTCCAGAAAGAGAAGATGGTGGC 168
Db	1 GCATTAGAAGATGCAGTATTGCACAGTATTACTTATTCCAAGAAAAATGSTATTA 60
Oy	169 AATGAAGGAGCATGTGCACGGCCGTTACAACACCAACACCCGGTTCTTAAAGGCTCTGT 228
Db	61 AATGAAGAACACAGTAGGAACACGCTTTACAACATACTACACCGTTCAAGTGTTCAAGTT 120
Oy	229 GCTAAGCGTGGCTCCGGTGGGCTCTGTGGCCCTCTGGGGGTTCCGTGCGCTCCGGGGCAC 288
Db	121 ACTTCAGCTGGTGCAGTACTTCAGGTGGTTCAGTTACTTACGTTTCAGTTCAAGTTCCTCA 180
Oy	289 GTMGCATAGAGTGCGTCAGTGGCAAGCGGGGTTCCGGGAACAGTCGAAGAACAATCCA 348
Db	181 GTTGGTTAAGTTGCTTCAGTTGCTTCACAGTGGTTCAGTTAATTCAGACGTTCAATCTT 240
Oy	349 TCCTGACACTAGAGATTCCGACGCCAAGTCCTACGGCAGCACTCAACACACCGAGTTGACA 408
Db	241 TCAGATAATTAAAGGATTCCAGATCTTAATCTTACCGTGAATTTAAAACATRAGAGTTCAA 300
Oy	409 AACATATCTCTACATATCCAGGAGCTGAAGTACCCACAGTGTTCGACCTCACTAATCAT 468
Db	301 AATTACTGTTCACATATTAAAGAACTCAAAATATCCGAACTCTTGATTTAACCAATCAT 360
Oy	469 ATGCTGACACTGTGTGATTAACATTCATGGCTTCAAAATATCTGATTGACGGGTTACGAAG 528
Db	361 ATGTAACTTTGTGTGAATAATATCATGTTTTCAAAATTTTAAATTGATGGATATGAAGAA 420
Oy	529 ATCAATGAAGTCGTACAAAGTGAATTTCTACTTCGCACTGGTCGTGAAGGCCAAATGAT 588
Db	421 ATTATGATTAATTAATTAATAATTAACCTTTATTTATTTGATTTATTAAGACAATAATTAAT 480

OY	589	GACGTTTGGCGCAATGACTATTTCGAATTCCTCATCTTGAGATCATGAGCCAAAGC	648
Dd	481	GATGTATGTGCTAATGATTATTGTCCAATPACCTTYCAATCTTAATAATTCGTGCAAATGAA	540
OY	649	TTGACACTATTGAACAATTTGGTCTTCGGATATFCGCAAGCCCTCGACAACATCAAGAC	708
Dd	541	TTAGACGACTTAAAAAACTGTGTTCGGATATACAAAACCATTGACATATTTAAAGT	600
OY	709	AATGTGGAGAANGAAGGAAATTTATTTAAAAAGATTAAAGAAGACCATCGAGAACATTAAAC	768
Dd	601	AATGTAGGAAAAAATGSAAGATTACTATTAAAAAATTAACCAACCATAGCAATATTAAT	660
OY	769	GAGCTGATCCGAAGATCCAAAAAGGCCATAGACAAAAATTAAGATGCAACCAAGAGAA	828
Dd	661	GAATTATTTTGAAGGAAGTAGAAAAACAATTTGATCATAAATAAGAAATGAGATATGAAGA	720
OY	829	GAATAAGAAGATTGTGTACAGGCGCAGTACGACCTGTCTCATCTATTAACCAACAGCTTGA	888
Dd	721	GGAAAAAATAATTTATACCAAGCTCAATATATCTTTTATTTTACAAATPAACAAATTACAA	780
OY	889	GAAGCCCATTAACCTTCATCAGCGTACTGGAGAAGCCGATAGACACCCTCAAGAAGANTGA	948
Dd	781	GAAGCACATTAATTTTAAATTAAGCGTTTTAGAAAAACGTATTGACACTTTAAAAAATAAGAA	840
OY	949	AATATCAAAGAAGCTCTGACAAGATTATAGAAATTAAGATATCTCCGCCACGCCAATCT	1008
Dd	841	AACATTAAGAAATTAATCTGTATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	900
OY	1009	GGGAACACCCCTTAACACGCTGC---TGGAACAAGAACAAGAATAGAGAGACAGAGAAA	1065
Dd	901	GGAATATACCAAAATCCTCTCCCTGAGATTAAGAAAAAAGAAGTGGAGGACACGACAAGA	960
OY	1066	GAGATCAAGAAGANTGCCAAAAACCTTAAGTTCAACATAGAT	1107
Dd	961	AAATTAAGAAATTCGCAAAACCTTAATTAATTACATTGAT	1002

RESULT 58
PFMSALAB

```

LOCUS      PFMSALAB               1002 bp     DNA             linear   INV 22-MAY-2002
DEFINITION Plasmidium falciparum clone 807 major merozoite surface antigen
              (p190) gene, partial cds.
LOCUS      M7714                107714       nt             linear   uncloned
ACCESSION  M7714.2 GI:21040166
VERSION    T-cell epitope; major merozoite surface antigen; tripeptide repeat.
KEYWORDS   Plasmidium falciparum.
SOURCE     Plasmidium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ORGANISM   Plasmodium falciparum.
REFERENCE  1 (bases 1 to 1002)
AUTHORS    Jongwattives,S., Tanabe,K., Nakazawa,S., Yanaqi,T. and Kimbara,H.
TITLE      Sequence variation in the tripeptide repeats and T cell epitopes in
            p190 (MSA-1) of Plasmidium falciparum from field isolates
JOURNAL    Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
MEDLINE    92228018
PUBMED     1373473
REFERENCE  2 (bases 1 to 1002)
AUTHORS    Jongwattives,S.
TITLE      Direct Submissions
JOURNAL    Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn
            University, Rama4, Pathumwan, Bangkok 10330, Thailand
COMMENT    On May 21, 2002 this sequence version replaced gi:309696.
FEATURES
            location/Qualifiers
            1..1002
            /organism="Plasmodium falciparum"
            /db_xref="taxon:5833"
            /map="g"
            /clone="807"
            /sex="asexual"
            /dev_stage="erythrocytic"
            <1..>1002
            /gene="p190"
            <1..>1002

```

	BASE COUNT	420 a	126 c	153 g	303 t	
ORIGIN						
Query Match	9.8%;					Score 485.2; DB 3; Length 1002;
Best Local Similarity	68.5%;					Pred. No. 9.2e-96;
Matches 686; Conservative	0;					Mismatches 313; Indels 3; Gaps 1;
109	GGTTTGAAGATGCGCGTCTTACCGGATACAGCCGTTCCGGAAGAGACAGTGGTGTG	168				
Db	1 GCTTTAGAGATGCAAGTATTGACAGGTTATATGTTATTTCAAAAGAAAAATGTTATTA	60				
169	AATGAAGGAGACAGTGGCAAGCGCGCTTACCAACACAGACACCCGGTCTTAAAGGCTGTG	228				
Db	61 AATGAAGAACACAGTGGCAACAGCTGTTACCACTAGTATACACCGTTCAGTGGTTCAGT	120				
229	GCTACCGTGGTCCGGTGGGTCTGTGGCCCTTGGGGGTTCCGTCGCTCCGGCGGCGACG	288				
Db	121 ACTTCAGGTGGTTCAGTTACTTCAGGTGGTTCAGTTACTTCAGTTGCTTCAGTTGCTTCA	180				
289	GTGGCATCAGGTGGTCTCAGTGGCAAGCGCGGCTTCCGGGAACAGTGCAGGAACCAATCA	348				
Db	181 GTTGTTCAGTGGTCTCAGTGGTTCAGTGGTTCAGTGGTTCAGTGGTTCAGTGGTTCAGT	240				
349	TCTGCAACCTAGCGGATTCGAGCCCAAGTCTTACGCCGACCTCAGCACACCAGTGTGGA	408				
Db	241 TCAGTAAATTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	300				
409	AACATATCTCCATATCAGGAGGCTGAAGTACACACAGTGGTTCAGTGGTTCAGTGGTTCAGT	468				
Db	301 AATTTACTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	360				
469	ATGCTGACATGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	528				
Db	361 ATGTTAACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	420				
529	ATCATGAACCTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	588				
Db	421 ATTAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	480				
589	GACGTTTCGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	648				
Db	481 GAT	540				
649	TTGACGATTTGAAGAAGTGGTCTTGGGATATGCAAGCCTCTGACACATCAAGAC	708				
Db	541 TTGACGATTTGAAGAAGTGGTCTTGGGATATGCAAGCCTCTGACACATCAAGAC	600				
709	AATGGGGAAGATGGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	768				
Db	601 AATGATGGAAGATGGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	660				
769	GAGTGATGCAAGATTCACAAAAGACCATAGACAAAATTAAGATGACCAACAGAGGAA	828				
Db	661 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	720				
829	GAAAGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	888				
Db	721 GAAAGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	780				
889	GAAAGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	948				

c	83	54	1.1	3662	23	ABL05274
c	84	53.6	1.1	2943	21	AAAT0229
c	85	53.2	1.1	1167	23	AAAT0702
c	86	53.2	1.1	1413	23	AAAT1483
c	87	53.2	1.1	1413	23	AAAT85383
c	88	53.2	1.1	1413	23	AAAT90691
c	89	53.2	1.1	1606	23	AAAT69034
c	90	53.2	1.1	1822	23	AAAT72777
c	91	52.8	1.1	6644	20	AAAT3181
c	92	52.8	1.1	6767	22	AAAT6608
c	93	52.8	1.1	7372	20	AAAT33182
c	94	52.8	1.1	7797	20	AAAT33180
c	95	52.8	1.1	7996	20	AAAT33184
c	96	52.4	1.1	752	24	ABO14854
c	97	52.4	1.1	752	24	ABO14855
c	98	52.4	1.1	1080	24	ABT99537
c	99	52.4	1.1	4593	24	AAAT36070
c	100	52.2	1.1	1140	21	AAAT0071
c	101	52.2	1.1	1662	23	ABT05275
c	102	52	1.1	393	23	AAAT90684
c	103	52	1.1	9539	22	AAAT5347
c	104	52	1.1	9539	22	AAAT28180
c	105	51.6	1.0	863	22	AAAT94286
c	106	51.4	1.0	4590	22	AAAT4065
c	107	51.2	1.0	453	23	AAAT92079
c	108	51.2	1.0	484	24	AAAT21132
c	109	51.2	1.0	647	24	ABO39100
c	110	51.2	1.0	647	24	ABO39101
c	111	51.2	1.0	5454	21	AAAT0236
c	112	51	1.0	822	24	ABO44852
c	113	51	1.0	822	24	ABO44853
c	114	50.8	1.0	621	23	AAAT90688
c	115	50.8	1.0	10732	21	AAAT10594
c	116	50.6	1.0	8392	24	ABT33490
c	117	50.2	1.0	556	24	ABO36996
c	118	50.2	1.0	556	24	ABO36997
c	119	50.2	1.0	4000	18	AAAT91902
c	120	50	1.0	5661	19	AAAT21518
c	121	49.8	1.0	1563	24	ABT53204
c	122	49.8	1.0	2336	23	AAAT5662
c	123	49.6	1.0	441	23	AAAT5453
c	124	49.6	1.0	887	22	AAAT94064
c	125	49.6	1.0	1303	16	AAAT92525
c	126	49.6	1.0	7503	21	AAAT70206
c	127	49.6	1.0	49999	20	AAAT23895
c	128	49.6	1.0	50000	21	AAAT64139
c	129	49.4	1.0	693	24	ABO32104
c	130	49.4	1.0	693	24	ABO32105
c	131	49.4	1.0	867	20	AAAT91460
c	132	49.4	1.0	867	20	AAAT91461
c	133	49.4	1.0	867	22	AAAT42783
c	134	49.4	1.0	867	22	AAAT42784
c	135	49.4	1.0	1200	24	ABO39210
c	136	49.4	1.0	1200	24	ABO39211
c	137	49.4	1.0	1397	20	AAAT91462
c	138	49.4	1.0	1397	20	AAAT91463
c	139	49.4	1.0	1397	22	AAAT42785
c	140	49.4	1.0	1397	22	AAAT42786
c	141	49.4	1.0	7442	22	AAAT4686
c	142	49.4	1.0	8079	24	ABT92313
c	143	49.2	1.0	541	24	ABO45148
c	144	49.2	1.0	541	24	ABO45149
c	145	49.2	1.0	807	22	AAAT21312
c	146	49.2	1.0	875	22	AAAT95044
c	147	49.2	1.0	29392	19	AAAT15422
c	148	48.8	1.0	571	23	AAAT90731
c	149	48.8	1.0	506	23	AAAT68116
c	150	48.8	1.0	906	23	AAAT88069
c	151	48.8	1.0	993	23	AAAT68118
c	152	48.8	1.0	993	23	AAAT88067
c	153	48.8	1.0	993	23	AAAT90706
c	154	48.8	1.0	1149	21	AAAT51336
c	155	48.8	1.0	2579	23	ABT26856

Drosophila melanog
Plasmodium falcipar
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
Base sequence of t
Tumour suppressor
Base sequence of t
Cowpox virus bsr f
Base sequence of t
Oligonucleotide fo
Oligonucleotide fo
Mouse ischaemic co
Human liver glucoc
Arabidopsis thalia
Drosophila melanog
DNA encoding novel
Chemically pretrea
DNA transcription
Human neuroblastom
Yeast AOD9604-asso
DNA encoding novel
Consensus sequence
Oligonucleotide fo
Plasmodium falcipa
Oligonucleotide fo
Oligonucleotide fo
Gene encoding a su
Human immune syste
Oligonucleotide fo
Mannose-1-phosphat
Rattus norvegicus
Nucleotide sequenc
Drosophila melanog
DNA encoding novel
Human neuroblastom
Nicotiana glumabagl
Plasmodium falcipa
Murine LOBO homolo
Nucleotide sequenc
Oligonucleotide fo
Oligonucleotide fo
T. gondii immunoge
T. gondii nTc867 r
T. gondii cDNA enc
Reverse complement
Oligonucleotide fo
Oligonucleotide fo
T. gondii immunoge
T. gondii nTc1397
T. gondii cDNA enc
Reverse complement
Tumour suppressor
Chemically treated
Oligonucleotide fo
Oligonucleotide fo
Human breast cance
Human neuroblastom
Mouse poly Ig rece
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
Recombinant protei
Drosophila melanog

c	156	48.8	1.0	3231	23	ABL26858
c	157	48.8	1.0	8895	22	ABA06801
c	158	48.8	1.0	8895	22	ABA06801
c	159	48.8	1.0	8895	22	AAAT8888
c	160	48.8	1.0	8895	22	AAAT8888
c	161	48.8	1.0	9656	22	ABA05800
c	162	48.8	1.0	9656	22	ABA05800
c	163	48.8	1.0	9656	22	AAAT8884
c	164	48.8	1.0	9656	22	AAAT8884
c	165	48.8	1.0	14041	22	AAAT8024
c	166	48.8	1.0	16842	22	AAAT5412
c	167	48.8	1.0	16842	24	AAAT70384
c	168	48.8	1.0	16842	24	AAAT61336
c	169	48.8	1.0	16842	24	AAAT31419
c	170	48.4	1.0	792	18	AAAT25071
c	171	48.4	1.0	1612	7	AAAT60392
c	172	48.4	1.0	21354	22	AAAT6815
c	173	48.2	1.0	1690	16	AAAT92524
c	174	48.2	1.0	7536	24	ABT33464
c	175	48	1.0	727	24	ABO29312
c	176	48	1.0	727	24	ABO29313
c	177	48	1.0	2827	23	ABV23003
c	178	48	1.0	2827	23	ABV28839
c	179	47.8	1.0	390	23	AAAT5941
c	180	47.8	1.0	1684	23	AAAT90724

ALIGNMENTS

RESULT 1
AAAT21451
ID AAV21451 standard; DNA; 4940 BP.
XX
AC AAV21451;
XX
DT 23-SEP-1998 (first entry)
XX
DE P. falciparum modified gp190 DNA.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite; ss.
OS Plasmodium falciparum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 10..4929
FT /tag= a
FT /product= gp190
MO9814583-A2.
XX
PF 02-OCT-1997; 97WO-EP05441.
XX
PR 02-OCT-1996; 96DE-4040817.
XX
PA (BUA/) BUJARD H.
XX
PI Bujard H, Pan W, Tolle R;
XX
DR WPI: 1998-240088/21.
XX
DR P-PSDB; AAAT54145.
XX
PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX
PS Example 1; Fig 3c; 48pp; German.
XX
CC This sequence encodes a modified Plasmodium falciparum gp190/MSP-1

CC (merozoite surface protein) which has a reduced AT content resulting in
CC a higher stability of the protein. Such a protein is useful in
CC vaccines against malaria or for producing monoclonal antibodies (for
CC passive immunisation). The complete gp190 protein can now be produced
CC outside the parasite and has, at least over extended regions, the native
CC pattern of folding. Larger amounts of the protein can be produced
CC recombinantly than would be possible using the parasites as source.
XX

Sequence 4940 BP: 1669 A; 1156 C; 1060 G; 1055 T; 0 other;

Query Match 100.0%; Score 4938.4; DB 19; Length 4940;
Best Local Similarly 100.0%; Pred. No. 0;
Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Oy 1 CGCAGCGCTATGAAAAATCATTTCTCTCTGTCATTTCTGTTTATATCATCAATCT 60
Db 1 CGCAGCGCTATGAAAAATCATTTCTCTCTGTCATTTCTGTTTATATCATCAATCT 60
Oy 61 CAGTGCCTGACCCAGCAATCTATCAGAGCTGTTAAGAACTGGAAGCTTTGGAGAT 120
Db 61 CAGTGCCTGACCCAGCAATCTATCAGAGCTGTTAAGAACTGGAAGCTTTGGAGAT 120
Oy 121 GCCGCTCTACCGGATACAGCTGTCCAGAAAGAAAGATGCTCTAATGAGGACG 180
Db 121 GCCGCTCTACCGGATACAGCTGTCCAGAAAGAAAGATGCTCTAATGAGGACG 180
Oy 181 AGTGCAGGCGGCTTACAAACGACACCGGTTCTAAAGGCTGTCGCTAGCGGTGGC 240
Db 181 AGTGCAGGCGGCTTACAAACGACACCGGTTCTAAAGGCTGTCGCTAGCGGTGGC 240
Oy 241 TCCGGTGGGTCTGTGCGCTCTGCGGGGTTCCGTGCGCTCCGGCGGACGCTGGCATCAGGT 300
Db 241 TCCGGTGGGTCTGTGCGCTCTGCGGGGTTCCGTGCGCTCCGGCGGACGCTGGCATCAGGT 300
Oy 301 GGCTCAGTGGCAAGGGGGGTTCCGGGAAACAGTGAAGAACCAATCTGACACACT 360
Db 301 GGCTCAGTGGCAAGGGGGGTTCCGGGAAACAGTGAAGAACCAATCTGACACACT 360
Oy 361 AGCGATTCGACGCCAAGTCTTACGCCGACCTCAAGCACCGAGTGAAGAACTATTTCTTC 420
Db 361 AGCGATTCGACGCCAAGTCTTACGCCGACCTCAAGCACCGAGTGAAGAACTATTTCTTC 420
Oy 421 ACTATCAAGAGCTGAAGTACCCAGCTGTTCGACCTCACTAATCATATGCTGACACTG 480
Db 421 ACTATCAAGAGCTGAAGTACCCAGCTGTTCGACCTCACTAATCATATGCTGACACTG 480
Oy 481 TGTGTAACATTCATGCTTCAAAATATCTGATTTGACGGTTACGAAAGATCAATGAACCTC 540
Db 481 TGTGTAACATTCATGCTTCAAAATATCTGATTTGACGGTTACGAAAGATCAATGAACCTC 540
Oy 541 CTGTACAGTTGAATTTCTACTGACTGCTGAAGGGCAAACTGAAGAGTTTGGCGCC 600
Db 541 CTGTACAGTTGAATTTCTACTGACTGCTGAAGGGCAAACTGAAGAGTTTGGCGCC 600
Oy 601 AATGACTATTTGTCAAAATTCATTCAATTTGAAGATCAGAGCAAGATTGGACCTATTTG 660
Db 601 AATGACTATTTGTCAAAATTCATTCAATTTGAAGATCAGAGCAAGATTGGACCTATTTG 660
Oy 661 AAGAAAGTTGCTTCGGATATCGCAAGCCTCTCGAACATCAAGAGCAATGTGGGAAAG 720
Db 661 AAGAAAGTTGCTTCGGATATCGCAAGCCTCTCGAACATCAAGAGCAATGTGGGAAAG 720
Oy 721 ATGGAAGTTTATTTAAAAAATTAAGAAACCATCGAAGCAATTAAGAGCTGTATCGAA 780
Db 721 ATGGAAGTTTATTTAAAAAATTAAGAAACCATCGAAGCAATTAAGAGCTGTATCGAA 780
Oy 781 GAATTCAAAAAGACCATAGACAAAAATTAAGATGCAACCAAGAGAGAAAGAAAG 840
Db 781 GAATTCAAAAAGACCATAGACAAAAATTAAGATGCAACCAAGAGAGAAAGAAAG 840
Oy 841 TTGTACACAGGCCAGTACACCTGTCTCATATTAACAACAGCTTGAAGAGCCCATAAC 900
Db 841 TTGTACACAGGCCAGTACACCTGTCTCATATTAACAACAGCTTGAAGAGCCCATAAC 900
```

```
Oy 901 CTCATCAGGCTACTGAGAGGCGATAGACACCCCTCAAGAAATGAATAATCAAGAA 960
Db 901 CTCATCAGGCTACTGAGAGGCGATAGACACCCCTCAAGAAATGAATAATCAAGAA 960
Oy 961 CTGCTCGACAGATTAATGAATTAAGAAATCTCCGCCAGCCCACTGGGAAACCCCT 1020
Db 961 CTGCTCGACAGATTAATGAATTAAGAAATCTCCGCCAGCCCACTGGGAAACCCCT 1020
Oy 1021 AACACGCTGCTGACACAGACAAAGATTAAGAGACACGAGAAAGACATCAAGAGATC 1080
Db 1021 AACACGCTGCTGACACAGACAAAGATTAAGAGACACGAGAAAGACATCAAGAGATC 1080
Oy 1081 GCCAAACCATTAAGTTCAATAGATATCTGCTTACTGATGCCCTGAGCTGAGATAC 1140
Db 1081 GCCAAACCATTAAGTTCAATAGATATCTGCTTACTGATGCCCTGAGCTGAGATAC 1140
Oy 1141 TACTTGAGAGAGAAAGAAATTAAGAAATCTCCGCCAAAGTCGAGACAAAGAAATCA 1200
Db 1141 TACTTGAGAGAGAAAGAAATTAAGAAATCTCCGCCAAAGTCGAGACAAAGAAATCA 1200
Oy 1201 ACCGAACCTAATGAATATCCCAATGCTGACGTACCTCTGCTTATAAGATATCAAC 1260
Db 1201 ACCGAACCTAATGAATATCCCAATGCTGACGTACCTCTGCTTATAAGATATCAAC 1260
Oy 1261 AACGCTCTCAACGACCTCAATAGCTTGGTGATCTGATTAACCCCTTGATATACGAA 1320
Db 1261 AACGCTCTCAACGACCTCAATAGCTTGGTGATCTGATTAACCCCTTGATATACGAA 1320
Oy 1321 GAACCCCTTAAGAAATATCTTACACAGACAAATGAGAGAAAGATTATCAACGAAATCAAG 1380
Db 1321 GAACCCCTTAAGAAATATCTTACACAGACAAATGAGAGAAAGATTATCAACGAAATCAAG 1380
Oy 1381 GAGAAATCAAAATTAAGAAAGAAATTAAGAAATGAGAAAGTTAGGAAGACCGC 1440
Db 1381 GAGAAATCAAAATTAAGAAAGAAATTAAGAAATGAGAAAGTTAGGAAGACCGC 1440
Oy 1441 AGCAAAAGTCTAAACGATATCTAAGAGTATGAAAGCTGCTGAAGAGATCTATGAT 1500
Db 1441 AGCAAAAGTCTAAACGATATCTAAGAGTATGAAAGCTGCTGAAGAGATCTATGAT 1500
Oy 1501 TCCAAAATTCACAAATTAACATGACCTGACCAACTCGAGAAATGATGGAAACGCTAC 1560
Db 1501 TCCAAAATTCACAAATTAACATGACCTGACCAACTCGAGAAATGATGGAAACGCTAC 1560
Oy 1561 TCTTCAAAAGTGGAGAAACCTGACACACATATACCTTTGATCTATGAGAAATTCCTAAG 1620
Db 1561 TCTTCAAAAGTGGAGAAACCTGACACACATATACCTTTGATCTATGAGAAATTCCTAAG 1620
Oy 1621 CATATCTTGAAGAGCTGACCAAAAGCTCTTAAGTATGAGAGATTTCTCTGCGGAAC 1680
Db 1621 CATATCTTGAAGAGCTGACCAAAAGCTCTTAAGTATGAGAGATTTCTCTGCGGAAC 1680
Oy 1681 ATTGTTGTGAGAAAGAACTAAAGTATTACAAAGATCTCATTAAGTATGAGAAACGAG 1740
Db 1681 ATTGTTGTGAGAAAGAACTAAAGTATTACAAAGATCTCATTAAGTATGAGAAACGAG 1740
Oy 1741 ATCGAGACGCTTTGTTGAGACATTAAGAGATGAAGAAACAGTTGTTGAGAAAGATT 1800
Db 1741 ATCGAGACGCTTTGTTGAGACATTAAGAGATGAAGAAACAGTTGTTGAGAAAGATT 1800
Oy 1801 ACAAAGAGCAAAATTAACCGATGAGAAAGTCCGAGGAGCTCCGATATGTTTAAAGTC 1860
Db 1801 ACAAAGAGCAAAATTAACCGATGAGAAAGTCCGAGGAGCTCCGATATGTTTAAAGTC 1860
Oy 1861 CAAGTGCAAGAGTGTCTCTCATGAACAAAGATTAAGAAATCAAGAACTCAACTCAT 1920
Db 1861 CAAGTGCAAGAGTGTCTCTCATGAACAAAGATTAAGAAATCAAGAACTCAACTCAT 1920
Oy 1921 CTGAAGAAAGCTGAGTTTAAACATTAATATACATGTGCGGAATAGTTATTAACAGAGAAAT 1980
Db 1921 CTGAAGAAAGCTGAGTTTAAACATTAATATACATGTGCGGAATAGTTATTAACAGAGAAAT 1980
```

1981 AACGAGAACACTACTACTCTCATCTGTAATCAAGAAAGATAGACAACCTGAAAGTGTTC 2040
1981 AACGAGAACACTACTACTCTCATCTGTAATCAAGAAAGATAGACAACCTGAAAGTGTTC 2040
2041 ATGCCCCAAGTTCGAGAGCTGATCAACGAAAGAGAAAGAACTTAAACTGAAAGCAG 2100
2041 ATGCCCCAAGTTCGAGAGCTGATCAACGAAAGAGAAAGAACTTAAACTGAAAGCAG 2100
2101 TCAGATTAACCTCCGAGCTTCACAGAAAGAGATTAACGAGAGCTTACACCAACGCC 2160
2101 TCAGATTAACCTCCGAGCTTCACAGAAAGAGATTAACGAGAGCTTACACCAACGCC 2160
2161 GGACAAACAGGCGGTTCCAGCTTCGAAGGCGATAGCTGCAAGCTCAAGCAAGAGCAG 2220
2161 GGACAAACAGGCGGTTCCAGCTTCGAAGGCGATAGCTGCAAGCTCAAGCAAGAGCAG 2220
2221 AACGAGGCAAGGCTCCAGTCCAGTCCGAGTCCGAGGCTTAAAGCTCAAGTCCCTACA 2280
2221 AACGAGGCAAGGCTCCAGTCCAGTCCGAGTCCGAGGCTTAAAGCTCAAGTCCCTACA 2280
2281 CCACAGCTCTGTAATTAACAAGACCGAGAAATGTCAAGAACTGACTACTCTTGAAG 2340
2281 CCACAGCTCTGTAATTAACAAGACCGAGAAATGTCAAGAACTGACTACTCTTGAAG 2340
2281 CCACAGCTCTGTAATTAACAAGACCGAGAAATGTCAAGAACTGACTACTCTTGAAG 2340
2341 CTCTATGAGTTCCTGAATTAACATCTTACATCTGCAACAAATATATCTCTCTCAGC 2400
2341 CTCTATGAGTTCCTGAATTAACATCTTACATCTGCAACAAATATATCTCTCTCAGC 2400
2401 ACATGAGAGAGAAAGATTTCTTAAACAGTACAGATTAACCAAGGAGGAGAGAACTG 2460
2401 ACATGAGAGAGAAAGATTTCTTAAACAGTACAGATTAACCAAGGAGGAGAGAACTG 2460
2461 TCCTCTTGATGATCCACTGAGCTGCTGTTCAATATCCAGAAACAACTTCCGTTATAT 2520
2461 TCCTCTTGATGATCCACTGAGCTGCTGTTCAATATCCAGAAACAACTTCCGTTATAT 2520
2521 TCATATGTCATAGCCCTCAACAAATTCCTCTCTCACTGTTCAAGATATATGAAAG 2580
2521 TCATATGTCATAGCCCTCAACAAATTCCTCTCTCACTGTTCAAGATATATGAAAG 2580
2581 GAGATGTCGCAACCTGTTAAATCAAGACAAAGCAAGATTAAGAACTTCTTGAG 2640
2581 GAGATGTCGCAACCTGTTAAATCAAGACAAAGCAAGATTAAGAACTTCTTGAG 2640
2641 GAACTTAAGAAAGTCTTCACCTCTGTTAAACCTCTCTTCCAGCTTCAACCACTG 2700
2641 GAACTTAAGAAAGTCTTCACCTCTGTTAAACCTCTCTTCCAGCTTCAACCACTG 2700
2641 GAACTTAAGAAAGTCTTCACCTCTGTTAAACCTCTCTTCCAGCTTCAACCACTG 2700
2701 TCCTCTCACTTCAAGCAAGGCGGAGTGAAGGCTTACGACGACCTCTCAGTGCAC 2760
2701 TCCTCTCACTTCAAGCAAGGCGGAGTGAAGGCTTACGACGACCTCTCAGTGCAC 2760
2761 AACCTTAATTAACCTGTAAGTGTGGAAGATCTCTGCTCCGCAAGAAATGAAG 2820
2761 AACCTTAATTAACCTGTAAGTGTGGAAGATCTCTGCTCCGCAAGAAATGAAG 2820
2761 AACCTTAATTAACCTGTAAGTGTGGAAGATCTCTGCTCCGCAAGAAATGAAG 2820
2821 ATCTACCAAGAACTTATTTGACAGAAATGCTCCGAGAACTTCTACGAGAAATGAAG 2880
2821 ATCTACCAAGAACTTATTTGACAGAAATGCTCCGAGAACTTCTACGAGAAATGAAG 2880
2821 ATCTACCAAGAACTTATTTGACAGAAATGCTCCGAGAACTTCTACGAGAAATGAAG 2880
2881 GACAGGCAACATTTCTTAACGAGAGCTTCAAACTCTGGAATTTAAAGCGAGAT 2940
2881 GACAGGCAACATTTCTTAACGAGAGCTTCAAACTCTGGAATTTAAAGCGAGAT 2940
2941 ATCAACTCTCTTAACGAGTGAATTAACGTAAGAGCTGGAAGGAGCATCAATTAAGCTG 3000
2941 ATCAACTCTCTTAACGAGTGAATTAACGTAAGAGCTGGAAGGAGCATCAATTAAGCTG 3000
2941 ATCAACTCTCTTAACGAGTGAATTAACGTAAGAGCTGGAAGGAGCATCAATTAAGCTG 3000
3001 AAGAGAACTGCAACTGAGCTTGAAGCTTGAAGCAAGTGAAGAACTGGAAGTGAAG 3060
3001 AAGAGAACTGCAACTGAGCTTGAAGCTTGAAGCAAGTGAAGAACTGGAAGTGAAG 3060
3001 AAGAGAACTGCAACTGAGCTTGAAGCTTGAAGCAAGTGAAGAACTGGAAGTGAAG 3060
3061 CTCTTGAAGAAAGAGAGAGAGTGGCAAGTATTAAGATGAGATCAAGAACTGACTCTG 3120
3061 CTCTTGAAGAAAGAGAGAGAGTGGCAAGTATTAAGATGAGATCAAGAACTGACTCTG 3120

3061 CTCTTGAAGAAAGAGAGAGTGGCAAGTATTAAGATGAGATCAAGAACTGACTCTG 3120
3121 CTCGAGAGAGAGCTTGAAGAACTCAACTCACTGAGCAATCCGAAACAGCTAGCAG 3180
3121 CTCGAGAGAGAGCTTGAAGAACTCAACTCACTGAGCAATCCGAAACAGCTAGCAG 3180
3181 AACCTTCACTGTTCTTCAACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
3181 AACCTTCACTGTTCTTCAACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
3241 CTCGAG 3300
3241 CTCGAG 3300
3301 GAGCTTCTCTCTGAG 3360
3301 GAGCTTCTCTCTGAG 3360
3301 GAGCTTCTCTCTGAG 3360
3361 AGCTTCGAGAACTTCAAGTCTCTGTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 3420
3361 AGCTTCGAGAACTTCAAGTCTCTGTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 3420
3421 CTCGAG 3480
3421 CTCGAG 3480
3481 AACGAACTTAAAGAAAGAACTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3540
3481 AACGAACTTAAAGAAAGAACTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3540
3541 GTGAAATAGCAGCTGGAATTTTCAAGAGAGTCTCTGCTTGAAGAGAGAGAGAGAG 3600
3541 GTGAAATAGCAGCTGGAATTTTCAAGAGAGTCTCTGCTTGAAGAGAGAGAGAGAG 3600
3601 GTGAAATAGCAGCTGGAATTTTCAAGAGAGTCTCTGCTTGAAGAGAGAGAGAGAG 3660
3601 GTGAAATAGCAGCTGGAATTTTCAAGAGAGTCTCTGCTTGAAGAGAGAGAGAGAG 3660
3661 ACTGATGTCGAG 3720
3661 ACTGATGTCGAG 3720
3661 ACTGATGTCGAG 3720
3721 GATGAGCTCATATTTGCTTATCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
3721 GATGAGCTCATATTTGCTTATCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
3781 GAGTGTGTCACGAGTGAAGTGTGTCACCTCTCTGCTGATTAATTAATTTCTGCAAAATC 3840
3781 GAGTGTGTCACGAGTGAAGTGTGTCACCTCTCTGCTGATTAATTAATTTCTGCAAAATC 3840
3841 GAGTGTGTCACGAGTGAAGTGTGTCACCTCTCTGCTGATTAATTAATTTCTGCAAAATC 3900
3841 GAGTGTGTCACGAGTGAAGTGTGTCACCTCTCTGCTGATTAATTAATTTCTGCAAAATC 3900
3841 GAGTGTGTCACGAGTGAAGTGTGTCACCTCTCTGCTGATTAATTAATTTCTGCAAAATC 3900
3901 AAACAGCTGGAAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3960
3901 AAACAGCTGGAAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3960
3901 AAACAGCTGGAAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3960
3961 CGCTTAAATTAAG 4020
3961 CGCTTAAATTAAG 4020
3961 CGCTTAAATTAAG 4020
4021 GAGCTGAGCTCTCTTAACCTGTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
4021 GAGCTGAGCTCTCTTAACCTGTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
4021 GAGCTGAGCTCTCTTAACCTGTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
4081 AGGATTAATTTCTGCTTACTTCAACTATATCAAGAGAGAGAGAGAGAGAGAGAGAG 4140
4081 AGGATTAATTTCTGCTTACTTCAACTATATCAAGAGAGAGAGAGAGAGAGAGAGAG 4140
4141 TTGCTTAATATGTCGAGGAGATTTACAGATCTGAGGAGAGAGAGAGAGAGAGAGAG 4200
4141 TTGCTTAATATGTCGAGGAGATTTACAGATCTGAGGAGAGAGAGAGAGAGAGAGAG 4200

```
Db 4141 TTCCTAATGATGTCGTGGGATATTACAGATCCAGGAAATAACAGCTGACCTT 4200
Qy 4201 GACTCTATTAAAAATATACAGATTAAGGAGGAGGAATGAAAAATATCTGCCCTC 4260
Db 4201 GACTCTATTAAAAATATATACAGATTAAGGAGGAGGAATGAAAAATATCTGCCCTC 4260
Qy 4261 CTGATAACATGAAACCCCTGTACAGAGATGAAAGCAAAATCGACCTCTCTAATT 4320
Db 4261 CTGATAACATGAAACCCCTGTACAGAGATGAAAGCAAAATCGACCTCTCTAATT 4320
Qy 4321 CACCTGGAGGCCAAGTCTCTAATCTTACGAGAGACATGTGAAAGTTAAATC 4380
Db 4321 CACCTGGAGGCCAAGTCTCTAATCTTACGAGAGACATGTGAAAGTTAAATC 4380
Qy 4381 AAGAGAGTGAATCTCCCTCAAAACATCCAGACAGCGGAGATTTCCAGAAAAATAC 4440
Db 4381 AAGAGAGTGAATCTCCCTCAAAACATCCAGACAGCGGAGATTTCCAGAAAAATAC 4440
Qy 4441 AATTTGCTGGAATTCGAGACCTGTCTACGATTATTAACCAACAAATCTCTGACCAAG 4500
Db 4441 AATTTGCTGGAATTCGAGACCTGTCTACGATTATTAACCAACAAATCTCTGACCAAG 4500
Qy 4501 TTTCTGTCCACTGCGCATGTGTTGAAAACCTCGCCAAAACAGTCTGAGCAATCTGCTC 4560
Db 4501 TTTCTGTCCACTGCGCATGTGTTGAAAACCTCGCCAAAACAGTCTGAGCAATCTGCTC 4560
Qy 4561 GAGGGAACCTCGCAGGGGATGCTGAACATCTCCAGCAACATGGCTGAAGAAACAGTGC 4620
Db 4561 GAGGGAACCTCGCAGGGGATGCTGAACATCTCCAGCAACATGGCTGAAGAAACAGTGC 4620
Qy 4621 CCCCAGATAGCGGCTGTTTTCAGGCATCTGAGCAGCGCAGAGAGTGAAGTGTCTCTG 4680
Db 4621 CCCCAGATAGCGGCTGTTTTCAGGCATCTGAGCAGCGCAGAGAGTGAAGTGTCTCTG 4680
Qy 4681 AACTACAAACAGAGAGATTAAGTGGCTGAGAAACCCAAACCTTACCTGCAATGAAAC 4740
Db 4681 AACTACAAACAGAGAGATTAAGTGGCTGAGAAACCCAAACCTTACCTGCAATGAAAC 4740
Qy 4741 AATGCGGGGTGTGACGCCGATGCTAAATGACCGCAGAGAGACAGGGCTTAACGGAAG 4800
Db 4741 AATGCGGGGTGTGACGCCGATGCTAAATGACCGCAGAGAGACAGGGCTTAACGGAAG 4800
Qy 4801 AAAATCAGATCGAGTGTACTTAAGCCGACATCTATCCACTCTTGCAGCGGATTTTTTGC 4860
Db 4801 AAAATCAGATCGAGTGTACTTAAGCCGACATCTATCCACTCTTGCAGCGGATTTTTTGC 4860
Qy 4861 TCCAGCTCTAATTTCTCTGGGCATCTCTCTCTGATCTCATGCTGATCCTGTACAGC 4920
Db 4861 TCCAGCTCTAATTTCTCTGGGCATCTCTCTCTGATCTCATGCTGATCCTGTACAGC 4920
Qy 4921 TTCACTAATTAATGATCGATGG 4940
Db 4921 TTCACTAATTAATGATCGATGG 4940

RESULT 2
AAV35363
ID AAV35363 standard; DNA; 4940 BP.
XX
AC AAV35363;
XX
DT 23-SEP-1998 (first entry)
XX
DE P. falciparum gp190 DNA.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
XX monoclonal antibody; passive immunisation; parasite; ss.
XX OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 10..4929
FT /*tag= a
```

```
FT /product= gp190
PN W09814583-A2.
XX
PD 09-APR-1998.
XX
PF 02-OCT-1997; 97WO-EP05441.
XX
PR 02-OCT-1996; 96DE-4040817.
XX
PA (BUA/) BUJARD H.
PI Bujard H, Pan W, Tolle R;
DR WPI: 1998-240088/21.
P-PSDB: AAM54145.
XX
PT Recombinant production of complete gp190/MSP-1 plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX
PS Example 1; Fig 3c; 48bp; German.
XX
CC This sequence encodes the Plasmodium falciparum gp190/MSP-1 (merozoite
CC surface protein). This gene is used in a method for stabilising the
CC gene sequences by reducing the AT content. Such products are useful in
CC vaccines against malaria or for producing monoclonal antibodies (for
CC passive immunisation). The complete gp190 protein can now be produced
CC outside the parasite and has, at least over extended regions, the native
CC pattern of folding. Larger amounts of the protein can be produced
CC recombinantly than would be possible using the parasites as source.
XX
SQ Sequence 4940 BP; 2196 A; 597 C; 687 G; 1460 T; 0 other;

Query Match 57.4%; Score 2837.6; DB 19; Length 4940;
Best local similarity 73.4%; Pred. No. 0;
Matches 3626; Conservative 0; Mismatches 1314; Indels 0; Gaps 0;

Qy 1 CGACGCCGTAAGAAATCTTCTCTCTGTCATCTTCTGTTTATCATCAATCACT 60
Db 1 CGACGCCGTAAGAAATCTTCTTATGTTCAATCTCTTTTATTAATCAATCA 60
Qy 61 CAGTGCCTGACCCAGCAATCCATCAGAGCTGTTAAGAAACAGCTTGGAGAT 120
Db 61 CAATGTGAACACATGAAGATTAATCAAGAACTTGCAAAACAGTGAAGAT 120
Qy 121 GCCGCTTACCGGATACAGCCTGTTCCAGAGAGAAAGATGTGCTGAATGAAGCAG 180
Db 121 GCAGATATGACAGGTTATAGTTTATTCAAAAGAAAAATGTAATGAAGAAC 180
Qy 181 AGTGCACAGGCGCTTACACAGCACACCCGCTTAAAGGCTGTGCTAGCGGTGGC 240
Db 181 AGTGCACAGGCTGTACACAGTACACCTGCTTAAAGGCTGTGCTAGCGGTGGT 240
Qy 241 TCCGGTGGGTCTGTGGCCCTGTGGGGGTTCCGTCGCTCCGCGGAGCGTGGCATCAGT 300
Db 241 TCAGGTGGCTCAGTTGCTTCAAGTGGCTCAGTCTGCTTCAAGTGGCTCAGT 300
Qy 301 GGCTCAGTGGCAAGCGGGCTTCGGGGAACAGTGCAGAGAACCAATTCATGCAACTCT 360
Db 301 GGCTCAGTGGCTTCAAGTGGCTTCAAGTGGCTTCAAGTGGCTTCAAGTGGCT 360
Qy 361 AGCATTTCCGACGCCAAGTCTTACGCGCAGCTCAAGCAACGAGTGAAGAAATCTCTC 420
Db 361 AGTATTCAGATGCTTAAATCTTACGCTGATTTAAACACAGATGCAAAATTAAGTTA 420
Qy 421 ACTATCAAGAGAGCTGAAGTATCCACAGTGTGTCAGCTCACTAATCATATCTGACACTG 480
Db 421 ACTATCAAGAGAGCTGAAGTATCCACAGTGTGTCAGCTCACTAATCATATCTGACACTG 480
Qy 481 TGTGATAACATTCATGCTTCAAAATATCTGATGAGCGGTAGCAGAGATCAATGAACTC 540
Db 481 TGTGATAATATTCATGCTTCAAAATATCTGATGAGATGAGAAATTAATGAATTA 540
```

```
QY 541 CTGTACAGTGAATTTCTACTCGACTTGTGAAGGCCAACTGATGACGTTTGCC 600
  || || || || || || || || || || || || || || || || || || || ||
Db 541 TATATATAAATTAACTTTATTTGATTTATTAAGACAAATTAATGATGATGCT 600
QY 601 AATGACTATTGTCAAATTCCTCATTTGAGATCAGAGCCACAGATGGACGATG 660
  || || || || || || || || || || || || || || || || || || || ||
Db 601 AATGATTAATGTCAAATTTACTTCAATCTTAAATTCGTCCAATGATTAAGTACT 660
QY 661 AAGAGTGTGCTTGTGATTCGACAGCCCTGTGACACATCAAGGACATGGGGAA 720
  || || || || || || || || || || || || || || || || || || || ||
Db 661 AAAAAGCTGTGTGCGATTTAGAAAACCTTGAACAATTTTAAGCTAATGTAGAAA 720
QY 721 ATGGAAGATTTATTTAAAAAGATTAAGAACCCATCGAGAATCTTAAGCATGAT 780
  || || || || || || || || || || || || || || || || || || || ||
Db 721 ATGGAAGATTTATTTAAAAAATTTAAAAAATTTAAATGATTAATTTAA 780
QY 781 GAATCCAAAAAGCCATAGCAAAAATTAAGATTAAGCAACCAAGAGAAAGAAAG 840
  || || || || || || || || || || || || || || || || || || || ||
Db 781 GAAAGTAGAAGAAACAAATGATTAAGAAATTAAGAACTTAAGAAAGAAAAA 840
QY 841 TTGTACAGGCGCCAGTACGACTGTCCATCTATACAAACAGCTTGAAGAGCCAT 900
  || || || || || || || || || || || || || || || || || || || ||
Db 841 TTATACCAAGCTCAATATGATCTTTCTATTTACAAATTAAGAAAGACACAT 900
QY 901 CTCATCAGCGTACTGGAAGAGCGCATAGACCCCTCAAGAGATGAATAATCA 960
  || || || || || || || || || || || || || || || || || || || ||
Db 901 TTATTAAGCGCTTTTAAAGAAACGATTTGACCTTTAAAAAATGAAAAATTA 960
QY 961 CTGCTGACAGATTTATGAAATTAAGAAATCTCGCGACGCAAGCTGGAGAC 1020
  || || || || || || || || || || || || || || || || || || || ||
Db 961 TTACTTGATTAAGATTAATGAATTAATTAATCCCGACGCGCAATCTGGAAT 1020
QY 1021 AACACGCTGTGACAGAAACAAAGATTAAGAGAGACGAGAAAGAGATCAAG 1080
  || || || || || || || || || || || || || || || || || || || ||
Db 1021 AATACCTCTCTGATTAAGAAACAAAAAATGAGAGAACGAAAAAGAAATTA 1080
QY 1081 GCCAAACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140
  || || || || || || || || || || || || || || || || || || || ||
Db 1081 GCCAAACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140
QY 1141 TACTTGAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1200
  || || || || || || || || || || || || || || || || || || || ||
Db 1141 TATTTAAGAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 1200
QY 1201 ACCGAACCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260
  || || || || || || || || || || || || || || || || || || || ||
Db 1201 ACTGAACCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260
QY 1261 AACGCTCTCAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320
  || || || || || || || || || || || || || || || || || || || ||
Db 1261 AATGCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320
QY 1321 GAACCCCTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
  || || || || || || || || || || || || || || || || || || || ||
Db 1321 GAACCAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
QY 1381 GAGAGATCAAAATTAAGAGAAAGAAATTAAGAGTGAAGAAAGATTAAGAG 1440
  || || || || || || || || || || || || || || || || || || || ||
Db 1381 GAAAAAATTAATTAAGAAAAAATTAAGATTAAGATTAAGATTAAGATTAAG 1440
QY 1441 AGCAAAAGTCTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAT 1500
  || || || || || || || || || || || || || || || || || || || ||
Db 1441 TCTAAGCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
QY 1501 TCCAAATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1560
  || || || || || || || || || || || || || || || || || || || ||
Db 1501 AGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
QY 1561 TCTTAAAGAGTGAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 1620
  || || || || || || || || || || || || || || || || || || || ||
Db 1561 TCTAATTAAGAGTGAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 1620

QY 1621 CATTAATCTTGAGAGAGCTCAACCAAGCTTTAAGTATTAAGAGAGATTA 1680
  || || || || || || || || || || || || || || || || || || || ||
Db 1621 CATTAATCTTGAGAGAGCTCAACCAAGCTTTAAGTATTAAGAGAGATTA 1680
QY 1681 ATTGTTGTGAGAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 1740
  || || || || || || || || || || || || || || || || || || || ||
Db 1681 ATAGTGTGAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 1740
QY 1741 ATGAGAGAGCTTTGTGAGAGATTAAGAGAGATTAAGAGATTAAGAGAT 1800
  || || || || || || || || || || || || || || || || || || || ||
Db 1741 ATGAGAGAGATTAAGAGATTAAGAGAGATTAAGAGATTAAGAGATTA 1800
QY 1801 ACAAAAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1860
  || || || || || || || || || || || || || || || || || || || ||
Db 1801 ACTTAAAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1860
QY 1861 CAAGTGTGAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 1920
  || || || || || || || || || || || || || || || || || || || ||
Db 1861 CAAGTGTGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1920
QY 1921 CTGAGAGAGCTGTGAGAGATTAAGAGATTAAGAGATTAAGAGAGAT 1980
  || || || || || || || || || || || || || || || || || || || ||
Db 1921 TTAAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 1980
QY 1981 AACGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 2040
  || || || || || || || || || || || || || || || || || || || ||
Db 1981 AACGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 2040
QY 2041 ATGCCCAAGCTGTGAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2100
  || || || || || || || || || || || || || || || || || || || ||
Db 2041 ATGCCCAAGCTGTGAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2100
QY 2101 TCAGATTAAGCTGTGAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2160
  || || || || || || || || || || || || || || || || || || || ||
Db 2101 TCAGATTAAGCTGTGAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2160
QY 2161 GAGAGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2220
  || || || || || || || || || || || || || || || || || || || ||
Db 2161 GAGAGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2220
QY 2221 AACGAGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2280
  || || || || || || || || || || || || || || || || || || || ||
Db 2221 AACGAGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2280
QY 2281 CCAGAGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2340
  || || || || || || || || || || || || || || || || || || || ||
Db 2281 CCAGAGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2340
QY 2341 CTGATGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2400
  || || || || || || || || || || || || || || || || || || || ||
Db 2341 CTGATGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2400
QY 2401 ACTATGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2460
  || || || || || || || || || || || || || || || || || || || ||
Db 2401 ACTATGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2460
QY 2461 TCTGATGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2520
  || || || || || || || || || || || || || || || || || || || ||
Db 2461 TCTGATGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2520
QY 2521 TCTATGATGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2580
  || || || || || || || || || || || || || || || || || || || ||
Db 2521 TCTATGATGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2580
QY 2581 GAGATGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2640
  || || || || || || || || || || || || || || || || || || || ||
Db 2581 GAGATGAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2640
QY 2641 GAAGCTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2700
  || || || || || || || || || || || || || || || || || || || ||
Db 2641 GAAGCTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2700
QY 2701 TCTCTGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAG 2760
  || || || || || || || || || || || || || || || || || || || ||
```

Db 2701 TCATTAAACCTCCAGAGATAAACCCGAAGTAGTGCAATGATGATCATCTTCTACA 2760
Qy 2761 AACCTTAATACTACGTGAACCTGTTTGAGAACATCTGCTCTCGCGAGAAATAGAAC 2820
Db 2761 AATTGAATATAGTGTAAAAATTATTTGAAAAAATCTGAGCTCTGAAAAAACAATAAT 2820
Qy 2821 ATCTAACGAGAACTTATGAGACAGAAATCGTCCGAATCTCTACGAAAGATACTGAAA 2880
Db 2821 ATATACCAAGAAATTAATGATGCTAAAAAGTAGTGAACCTTTAGAAAAGATATATAAA 2880
Qy 2881 GACAGCGACATCTCTATTAAGAGAGAGCTTCACTTAATCTGTAATCTAAAGCCGAT 2940
Db 2881 GATAGTGATACATTTTATTAAGAACTTTTACAAATTTTGTAAATCTTAAAGCTATGAT 2940
Qy 2941 ATCAACTCTCTTAACGATGAATCTTAACGTAAGAAAGCTGAAGAGACATCAATAGCTG 3000
Db 2941 ATTAATCTATTAAGATGAATCAAAAAGAAATTAAGAAAGATATTAATTAATTA 3000
Qy 3001 AAGAAAGACCTGCAACTGAGCTTGACCTGTACAAACAGTCAAACTGAAGAGAGA 3060
Db 3001 AAAAAAATTTACAGTATATCAATTTGATTTATTAATTAATTAATTAATTAATTAAGAGA 3060
Qy 3061 CTCTTCGACAAGAAAGACAGCTCGCAAGTATTAAGATGCAATCAAGAGTGAAGCTG 3120
Db 3061 TTATTTGATTAAGAAAGAAAGAGTGTGTAATATTAATTAATTAATTAATTAATTAATTA 3120
Qy 3121 CTCAAGGACAGCTTGAAGAACAACTCACTCACTGAACAAATCGAAACAGTACTGAG 3180
Db 3121 TTAAAGAAACATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3180
Qy 3181 AACTTCAGTGTCTTCAACAAAGAAAGAAAGAGCCGAGATCGCCGAGACAGAAACCT 3240
Db 3181 AACTTCCTGTCTTCTTAAACAAAAAAGAGAGCTGAATAGCAGAACTGAAGAAACACA 3240
Qy 3241 CTGAGAAACCAAGATTCCTCTCAACACTACAAAGAGCTCTCAAGTATTAATAGC 3300
Db 3241 TTAGAAACACAAAAATATTAATTAAGACATTAATTAAGACATTTGAATTAATTAATG 3300
Qy 3301 GAGCTTCTCTCTGAGAGACTCTCTCGAGAGAGACATCCAGCCGAGATTAATAGCC 3360
Db 3301 GAATCATCTCATTAATAAAAACTTAAGTAGAAGATCAATTAACCAAGAAATTAATTAAT 3360
Qy 3361 AGCCTCGAGAACTTCAGAGCTCTCTTAAGCTCGAAAGCAGAGCTGAAGAGCAACTGAG 3420
Db 3361 AGTTAGAAAACTTTAAAGCTTTAAGTAAATTAAGAGAAAAATTAAGAGATTAATTAAT 3420
Qy 3421 CTGAGAAAGAAAGAGCTGAGCTCTCTAGCGAGCTGATCACTGAGCTGCGAGCTC 3480
Db 3421 TTAGAAAAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3480
Qy 3481 AAGGAAGCTATTAAGAGACAAAGATACACCGGCAATAGCCCAAGCGAGATTAATAGAG 3540
Db 3481 AAGGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540
Qy 3541 GTGATAAGAGAGCTGAGATTTTACAAAGATTTCTGCTGAGAGAAAGAGATGTCGCACT 3600
Db 3541 GTTAACAATGATTAAGATTTTACAAAAAATTTCTCCGAGAGAGAGATGTTTCAACA 3600
Qy 3601 GTGTGCTGATGTGCTGCGACACACTGAGAGCTGCAACTTAAGAGAGCTGATCT 3660
Db 3601 GTTGTGAAGTAGAAGTAGATCGACACATTAAGAACAAAGTCAACAAAGAACACATCA 3660
Qy 3661 ACTCATGTGAGAGCGGAGATCCAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3720
Db 3661 ACTCATGTGAGAGAGAGATCTTAACCAATTAATTAATTAATTAATTAATTAATTAAT 3720
Qy 3721 GATGACGATCATGTGCTGCTGCGAGAGAGAGAGAGAGATGATGATGATGATGATGATG 3780
Db 3721 GATGACGATCATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780
Qy 3781 CAGGTGCTACCGGTGAGGCTGCTCACTCTCCGATTTGATTAACATTTCTGCAAAATC 3840
Db 3781 CAGGTGCTACCGGTGAGGCTGCTCACTCTCCGATTTGATTTTCTGCAAAATC 3840

Db 3781 CAAGTAGTAACGAGAGAAAGAGTAACTCTCTCCGTAATTTGATTAACATCTTTCAAAAT 3840
Qy 3841 GAGAAAGATTAACGAGAGTCTATCTGAACCTCTGCGAGAGCGTATAGGCTCTCAAG 3900
Db 3841 GAAAAATGATATGAGGTTTTATTTTAAACCAATTTAGAGAGGTTTTTAAGAGTTTTAA 3900
Qy 3901 AAACGCTGAGAAATTAAGCTGATGCTCTCAATGATCAAGCTGAAGAGAGATCTGCAAGC 3960
Db 3901 AAACATTAAGAAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 3960
Qy 3961 CGCTTAATTAAGAGAGAAATTTCAAGAAAGCTCTGAGAGAGAGATGATGATGATGATG 4020
Db 3961 CGATTTAATTAAGAGAGAAATTTCAAGAAAGCTTTAGATGATGATGATGATGATGATG 4020
Qy 4021 GAGCTGAGCTCTCTAATCTAGCTGTCAGAGACCATCAAGTTCTCAATTAAGAGAGAG 4080
Db 4021 GATTTAATCAATGAAGTAATTAATGATGTCAGAAATCCATTAATTTCTTAATTAAGAAA 4080
Qy 4081 AGGATTAATTTCTGCTAGTTTACAACTATATCAAGAGAGCTCCATGACAGGATTAAT 4140
Db 4081 AGAGATTAATTTCTTAGAGATTAATTAATTAATTAAGATTAATTAATTAATTAATTAAT 4140
Qy 4141 TTGCTAATGATGCTGCTGAGGATTTACAAAGATCTGAGCGAGAAATTAACAGTGAAGCT 4200
Db 4141 TTGCAATTAATGATGCTGCTGAGATTTATTAATTAATTAATTAATTAATTAATTAATTA 4200
Qy 4201 GAGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4260
Db 4201 GATTCATTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4260
Qy 4261 CTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4320
Db 4261 TTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4320
Qy 4321 CAGCTGAGAGCGCAAGGCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4380
Db 4321 CATTAGAGAGCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4380
Qy 4381 AAGGAGCTGAAGTACTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4440
Db 4381 AAGGAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4440
Qy 4441 AATTTCTGAGAGAGCTGAGAGCTGCTCAAGATTAATTAATTAATTAATTAATTAATTA 4500
Db 4441 AATTTCTGAGAGAGAGCTGAGAGCTGCTCAAGATTAATTAATTAATTAATTAATTAAT 4500
Qy 4501 TTTCTGCTCACTGAGAGGATTTGAAAAAAGCTGCGCAAAAGAGCTGAGCAATCTGCTC 4560
Db 4501 TTCTTGAAGAGAGGATTTGAAAAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4560
Qy 4561 GAGGCAAGCTGAGAGGAGCTGAGAGCTGCTGAGAGCTGCTGAGAGAGAGAGAGAGAGAG 4620
Db 4561 GATGGAAGCTGAGAGGAGAGATTTAAACATTTTCAACACACAGATGCTAAAAAACAAGT 4620
Qy 4621 CCCCAGATTAAGAGGCTGTTTCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680
Db 4621 CCACAAATTTCTGAGATGTTTCAAGAGATTTAAGATGAAGAGAGAGAGAGAGAGAGAG 4680
Qy 4681 AACTTCAAAAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740
Db 4681 AATTTCAAAAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740
Qy 4741 AATGCGGAG 4800
Db 4741 AATGCGGAG 4800
Qy 4801 AAAAAATCAAG 4860
Db 4801 AAAAAATCAAG 4860
Qy 4861 TCCAGCTCTAATTTCTGAG 4920
Db 4861 AGTTCTCTCACTCTTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4920

Query Match	Best Local Similarity	Matches 3587; Conservative	Score 54.8%;	Pred. No. 0;	Mismatches 1343; Indels 45; Gaps 3,
10	ATGAAATATATTTCTCTCTCTGTCATTTCTGTTTATATCATCAATATCTAGTCCG		69		
216	ATGAAATATATTTCTCTCTCTGTCATTTCTGTTTATATCATCAATATCTAGTCCG		275		
70	ATGAAATATATTTCTCTCTCTGTCATTTCTGTTTATATCATCAATATCTAGTCCG		129		
276	ATGAAATATATTTCTCTCTCTGTCATTTCTGTTTATATCATCAATATCTAGTCCG		335		
130	ATGAAATATATTTCTCTCTCTGTCATTTCTGTTTATATCATCAATATCTAGTCCG		189		
336	ATGAAATATATTTCTCTCTCTGTCATTTCTGTTTATATCATCAATATCTAGTCCG		395		
190	ATGAAATATATTTCTCTCTCTGTCATTTCTGTTTATATCATCAATATCTAGTCCG		249		

Db	396	GCTGTACAACTATGATACACCTGGTTCAAAGGGTTGAGTTGCTTCAGGTGGGTTCAAGTGGC	455
Qy	250	TCTGTGGCTCTCTGGGGGTTCCGTCGCCCTCCGGCGGACGGTGGCATCAGGTGGCTCACTG	309
Db	456	TCAGTTCGCTCAGGTGGCTCAGTTGGCTTCAGGGGCTCGAGTTCAGTGGGCTCACTT	515
Qy	310	GCAAGCGGGCTTCCGGGAACAGTCGAAAGAACATCCATCGACACCTAGACACTCC	369
Db	516	GCTTCAGGTGGTTAGGTAAATTACAGACGTACAAATCTTCAGATTAATTCAGATGATTTCA	575
Qy	370	GACGCCAAGTCTACGGCGACCTCAAGCACCGAGTGAAGAACTATCTCCTCAATCAAG	429
Db	576	GATGCTAAATCTTCGCTGATTTTAAACACAGAGTACGAATTAATCTGTTAACTCAAA	635
Qy	430	CAGCTGAAGTACCCACAGTTGTCGACATCATAATCTGACACTGTGTGATTAAC	489
Db	636	GAACCTCAATATCCTCAACCTTGTGATTTAACTAAATCAATGTTAACCTTGTGTGATAT	695
Qy	490	ATTCATGCTTCAAAATATCTGATTTGATGAGGTTACGAGAGATCAATGAACCTCCTGCAAG	549
Db	686	ATTCATGCTTCAAAATTTTAATTTGATGATATGAAGAAATTAATGAATTAATATATATAA	755
Qy	550	TTGAATTTCTACCTTGACCTGCTAAGGGCCAAACTGAATGACGTTGGCCCAATGACTAT	609
Db	756	TTAAACCTTTATTTTGATTTATTAAGGCAAAATTAATGATGTATGTCATATGATATAT	815
Qy	610	TGTCAAATTCATTTCAATTTTGAAGATCAGAGCCACAGCATTTGGACGTATTTGAAGAAGTTG	669
Db	816	TGTCAAATACCTTTCAATCTTAAATTTGCGCAAAATGAATTTAGACGTACTTAAAAAATT	875
Qy	670	GTCCTTCGATTTGCCAAGCCCTCGACCAATCAAGCAATGATGGGAAAGTGAAGAT	729
Db	876	GTTTCGGATATAGAAAACCATTTAGCAATATTTAAAGATTAATGTAGAAAAATGCAAGAT	935
Qy	730	TATATTAATAAAGATTAAGAAGACCATCGAATCAATTAACGAGCTGATCGAAGATCCAAA	789
Db	936	TACATTAATAAATAATTAATAAACCATTAAGAAATATTAATGAATTTAATGAAGAAAGTAAAG	995
Qy	790	AAGCATATGACAAATAATTAAGATGCAACCAAGAGAGAAAGAGAGAAGTTGATCCAG	849
Db	996	AAAACAATTTGATTAATAAATAAGAAATGCACTAAAGAAAGAAAGAAAAAATAATTAATCCAA	1055
Qy	850	GCCCGATGACGCTGTCATCTTAATCAAAACACTTGAAGAAAGCCATTAACCTCATCTACG	909
Db	1056	GCTCAATATGTATCTTCTTATTTACAAATTAACATTTGAAGAAAGCCATTAATTAATTAAGC	1115
Qy	910	GTACTGGAAGCGCATATGACACCCCTCAGAGAAATGAAGAAATATCAAGAACTGCTCGAC	969
Db	1116	GTTTATAGAAAACGTATTTGACACTTTTAAAAAATAATGAAGAACTTAAGSAATTACTGAT	1175
Qy	970	AAGATTAATGAATTAAGAATCCTCGCGCACCACTGTGGAGACACCCCTATACACGTG	1029
Db	1176	AAGATTAATGAATTAATAAATCCGCCACCGCCAAATTCGAGAAATACACAAATCTCTC	1235
Qy	1030	CTGGAACAAGACAAGATATAGAGAGGACAGAAAGAGATCAAAAGAGATGCCCAAAAC	1089
Db	1236	CTTGATTAAGCAAAAAATTCGAGGAACAGAAAAAGAAATTAAGAAATTTGCCAAAACCT	1295
Qy	1090	ATTAAGTCAACATAGATTTCTCTTCTTACTGATCCCTTGAGCTGGAGTACTACTTGTGA	1149
Db	1296	ATTAAATTTAATATTTGATAGTTTATTTACTGATCCACTTGAAATTTAGSAATATCATTTAAGA	1355
Qy	1150	GAGAAGATAAGAAATATAGACATCTCCGCAAAAGTCGAGACAAAGGAATCAACGACACT	1209
Db	1356	GAATAAATAAATAATTTGATATTAAGTCAAAAGTGAACAAAGGAATCAACATCAACCC	1415
Qy	1210	AATGAATATCCCATGTGTGTGAGTACCCCTCTCTTATTAAGATATCAACAACCTCTC	1269
Db	1416	AATGAATATCCAAATGAGATTTACTTATCTTCTTATTAAGATTAAGATTAATTAACCTTTA	1475
Qy	1270	AACGAGTCATTAATCTTCGGAGCTTATTAATCAACCCCTTGATATTAACGAAGAACCTCT	1329

Db 1476 AATGAACTTAATCTTTGGTGATTTAATTAATCCATTTGATTTATACAAAAGACCAAGT 1535
Qy 1330 AAGAAATCTACACAGAAATGAGAAAGAGTTATCAACGAATCAAGAGAGATC 1389
Db 1536 AAAACATATATACGATTAAGAAAGAAAAATTCATTAAGAAATTAAGAAAAATTT 1595
Qy 1390 AAAATGGAAGAAAGAAATTTAGAGTACAAAGAAAGTACGAGACCGAGCAAAAGT 1449
Db 1596 AAAATGAAAAAAAATTTGATCTGATAAAAATCTTACGAAAGACAGATCTAAGTCT 1655
Qy 1450 CTAACGATATCATTAAGAGATGAGAAAGCTGTGAACGAGATCTATGATCCAAATTC 1509
Db 1656 TTAATGATATTAACAAAGAAATATGAAAAATTAATTAAGAAATTTATGATAGCAAAATTC 1715
Qy 1510 AACATTAACATCGACCTGACCAACTGAGAAATATGATGGGAAAAAGCTACTTACAA 1569
Db 1716 AATTAATTAATATGATTTAATCTAATTTGCAAAAAATGATGGGTAAGAAAGATTTCAATTA 1775
Qy 1570 GTGAGAAACTGACACACCATTAATACCTTTGATCTGATGAAATTTCAAGCATATCTT 1629
Db 1776 GTTGAGAAACTTACACACCCCTAATCTTTGATCTGATCTGAAATTTCAACATATCTT 1835
Qy 1630 GAGAAAGCTACCAAGCTTTAAGTATATGAGAGCTATCTCTCGGACATTTGTGTG 1689
Db 1836 GAAAAATTACAAAAAGCTTTAAATATATGGAAGATTAATCTTTAAGGAATATAGTAGTT 1895
Qy 1690 GAGAAAGAACTAAGATTAATTAACAAGATCATTAAGATGAAAGAGAGATGAGAG 1749
Db 1896 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1955
Qy 1750 CTTGTTGAGAACTTTAAGAGATGAGAAAGCTTTGTTGAGAAAGATTAACAAAGAC 1809
Db 1956 TTAGTTGAAATATTTAAAAAGATGAGAAAGCTTTTGGAAAAAAAATTTCTAAAGAC 2015
Qy 1810 GAAATTAACAGATGAGAAAGATCTCGAGAGCTCCGATATTTGTTAAAGTCCAAAGTCAG 1869
Db 2016 GAAAAATTAACAGATGAGAAAGATTTTAAAGATATCTGACATTTGTAAGTAAGTAAGTTTCAA 2075
Qy 1870 AAGGTCTCTCATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1929
Db 2076 AAAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2135
Qy 1930 GTGAGATTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1989
Db 2136 GTAGAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2195
Qy 1990 CCATCTACCTCATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 2049
Db 2196 CCTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2255
Qy 2050 GTGAGAGCCTGATCAACGAGAGAGAAAGACATTAAGTGAAGAGACAGTCAAGTAAAC 2109
Db 2256 GTAGAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2315
Qy 2110 TCCGAGCCTTCCACAGAGAGAGATTAACCGACAGGCTACCCAGACAGCCGAGACAG 2169
Db 2316 TCGGAACCATCAACGAGAGAGAAATTAACAGGACAAAGCACTACAAACCTGACAAACAA 2375
Qy 2170 GCCGTTACGCTCTGAGAGGGGATAGCTGCAAGCTCAAGCAAGACAGACAGAGAGAGCA 2229
Db 2376 GCAGAGTCTGCTTTAAGAGGAGATTCAAGTACAAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2435
Qy 2230 CAGCCTCAGTGCAGTCCGCTTCCAGAGGCTAAAGCTCAAGTCCCTACACACACACT 2289
Db 2436 CAACCAACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2495
Qy 2290 CCTGTGAATTAACAGACAGAGAGATGCAAACTGACATCTGAGAAAGCTCTATGAG 2349
Db 2496 CCAGTAATTAATTAATAAGAAATGTTCCAAATTAAGATTAATCTGGAATAATTAATTAATCAA 2555
Qy 2350 TTCCGAATTAATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 2409
Db 2556 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2615

Qy 2410 GGAGGATTTCTTAACAGTAAAGATTAACCAAGAGAGAGAGATTAACCTGCTCTGT 2469
Db 2616 GAAAAATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2675
Qy 2470 GATCCACTGAGACCTGCTCTTCAATATCCAGAACATCTCCGTTATGATCTATCTATGTTTC 2529
Db 2676 GATCCATTAAGCTTATGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2735
Qy 2530 GATAGCTCAACA-----ATTCT 2547
Db 2736 GATAGCTTAACCATAGTTTATTCACAACTATTTATGTTTATGAAAAAATTTGTTAT 2795
Qy 2548 CTCTCAACCTGTCATGAGATATATGAGAGAGAGATGCTGACCTGATTAATCTC 2607
Db 2796 GTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2855
Qy 2608 AAAGCAACGACAGATTAAGAACTCTGAGAGAGCTAAGAAAGTCTCCACTCTGT 2667
Db 2856 AAGATTAATGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2915
Qy 2668 AAAACTCTCTCTCCAGCTCCATGACACCTGCTCTCACACCTCAAGACAGACGCCGA 2727
Db 2916 AAAACTCTCTCAAGTTCAATGACACCATTAATTAATTAATTAATTAATTAATTAATTAAT 2975
Qy 2728 GTGAGCGTAAAGACGACACCTCTCACTGACCAACCTTAATTAATTAATTAATTAATTAAT 2787
Db 2976 GTAACTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3035
Qy 2788 GAGAACATCTCTCTCTCGGCAAGAAATTAAGAAATCTACCAAGAACTTAATGAGACAGAA 2847
Db 3036 GAAACATTAATGAGCTGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3095
Qy 2848 TCGTCCGAGAACTTACGAGAAAGTACTGAAAGACACGACACCTCTATTAACGAGAGC 2907
Db 3096 ACTAGTGAACCTTTATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3155
Qy 2908 TTCACTAATCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2967
Db 3156 TTTACAAATTTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3215
Qy 2968 CGTAAGAGCTGAGAGAGACATCAATTAAGTGAAGAGACCTGCAACTGAGCTTGAC 3027
Db 3216 AGGAAGAAATTAAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3275
Qy 3028 CTGTACAACAGTAAACTGAAGTGAAGAGACCTCTGACANAGAAAGACAGTCCGC 3087
Db 3276 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3335
Qy 3088 AAGTAAAGATGACAGATCAAGAAAGTTGACTGCTCAAGAGACACTGAAGCAAACTC 3147
Db 3336 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3395
Qy 3148 AACTCACTGAACATCGAAGACAGTCTGAGAACTTCTGAGTCTTCAACAAGAG 3207
Db 3396 AATTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3455
Qy 3208 AAGGAAGCGAGATGCGGAGACAGAGAACTCTGAGAGAAACACCAATTTCTTCAAA 3267
Db 3456 AAAGAAAGCTGAATTAAGCAAGAACTGAACCAATTAAGAAACACCAATTAATTAATTAAGAA 3515
Qy 3268 CACTCAAGAGCCTGCTCAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3327
Db 3516 CATTTAAAGAGCTTTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3575
Qy 3328 GAGGAGACATTCAGACCGAGAGATTAAGTCAAGCTCGAAGCTTCAAGGCTCTGTCT 3387
Db 3576 GAAGAAATCAATTAACAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3635
Qy 3388 AAGCTCGAAGCAAGCTGAAGACAACTGAACTGGAAGAGAAAGAGCTCAAGCTACCTC 3447
Db 3636 AAATTAAGAGAAATTAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3695

OY	3448	TCCTAGCGAGCTCAATCCATCGATGGCGACCTCAAGGAATCATTTAAGAAACAAGAACTAC	3507
Dd	3656	TCMAAGSTTTACATCATTTTAATTTGCTGAATTTAAAGAAAGTATATAAAATATTAAT	37555
OY	3508	ACCGCAATACGCCCAAGCGAATAATATACAGCGTGAATTAAGCACTGGATCTTTACAG	3567
Dd	3756	ACAGGTATTTCTCCACGCGTAATAATATACGGATGTTAAACATATGCAATTAAGATCTTACAA	3815
OY	3568	AAAGTTCTGGCTGAAGAACAGATGTCCACTGTGGTGTCTGAATCTGGCTCGACACA	3627
Dd	3816	AAATTTCTCCGACGAGAGAACAGATGTTGCAACAGTTGTAAAGTAAAGTGATGCCACACA	3875
OY	3628	CTGGACGAGTCTCAACCTAAGAAAGCCGATCTACATCATGTGGAGCGAGTCCAAATACA	3687
Dd	3876	TTAGAACAAAGTCAACCAAGAAAGAACAGCACTCAACTCATGTAGAGACACAGTCTACACA	3935
OY	3688	ATTACACATATCTCAAGACGTCCGAGATGAGTGCATGATGATCATGTCATTTGGCTATCTTC	3747
Dd	3936	ATTAAACATCTACAAATATGTTCGATGATGATGAGTATGATGAGTATCATATCTACTATTT	3995
OY	3748	GGCGAGACGAGAGAGACTACGATGACCTTCGGCAGGTGGTCACCGGTGAGGCTGTCACT	3807
Dd	3996	GGAGATATCCGAAGAAAGATTTATGATGATTTAGCAAGTATGTAACAGAGAAACAGTACT	4055
OY	3808	CTTTCCGATTTGATATCATCTTGTCTCCAAATTCGAAACAGATACGAAGTGTCTATCTG	3867
Dd	4056	ACTTCCGTAATTTGAATCACTACTTTCTAAATTTGAAATTAAGAAATATGAGGTTTATATTTA	4115
OY	3868	AAACCTCTGAGAGCGTCTATAGTCTCTCAGAAACACCTGGAGATTAACGTGATGACC	3927
Dd	4116	AAACCTTTAGCAGGTGTTTATAGAAGTTTAAAGAAACATTTAGAAATATACGTTATGACA	4175
OY	3928	TTCAATGTCAAGTGAAGACATTTCTGAACAGCGCTTTAATTAAGAGAGAAATTTCAAG	3987
Dd	4176	TTTATATGTTATGTTAAGGATATTTTAAATTCACGATTTAATTAACGTAAGAAATTTCAA	4235
OY	3988	AAACGCTTGGAGACGCTGATTTCCCTATTAAGACCTGACCCGCTCACTACGATGCTG	4047
Dd	4236	AAATGTTTAGATCAGATTTTAATTTCCATTAATAAGATTTTAACATCAAGTATATATGTGTG	4295
OY	4048	AAGGACCATATCAAGTTCCTCAATTAAGAGAGAGGAGATTAATTTCTGTCTAGTTACAC	4107
Dd	4296	AAAGATCCATATTAATTTCTTAATTAAGAAAGAAAGATTAATTTCTTAAGCAGTTATAT	4355
OY	4108	TATATCAAGACTCCATCCGACACCGATATCATTTGCTTAATGATGTGCTGGGTAATTAC	4167
Dd	4356	TATATTAAGGATTAATAATGATAGTACGATATTAATTTGCAAAAGATGTCTTGGATATAT	4415
OY	4168	AAGATCCGACGAAAAATCAAGTGCAGCTGATCTATTAATAAAGTATATCAAGAT	4227
Dd	4416	AAATATTTATCCGAAAAATATAATCAGATTTAGATTCATTTAAAAAATATATACAGAC	4475
OY	4228	AAGCAAGCGAATGGAATAATATCTGCCCTTCTCAATTAACATCGAAACCTGTACAG	4287
Dd	4476	AAACAGGTGAATATGGAATATACCTTCCCTTTTAAACAATATTSAGACCTTATATATA	4535
OY	4288	ACAGTGAACGACAAATCGAAGCTTCTGTATATTCACCTGAGAGCCAGGCTCTCAAT	4347
Dd	4536	ACAGTTATATGATTAATAATTTGATTTTATTTGTAATTCATTTGAAGCAAAAGTCTTAATAT	4595
OY	4348	ACTTACGAGAGACCATGTGAGAAAGTTAAATTAAGAGACTGACTACTCTCAAAACATC	4407
Dd	4596	ACATATGGAATATCAACCGTAAGAAATTAATAAGAACTTAATTAATCTTAATAACAAT	4655
OY	4408	CAAGCAAGCTGAGAGATTTTCAAGAAAAATTAACATTTGCTGGCAATTTGCACACCTGCT	4467
Dd	4556	CAAGCAAAATTTGGCAGATTTTAAAAAAATTAACAAATTTGCTTGGATTTGCTGATTTATCA	4715
OY	4468	ACCGATTAATACCAACAATCTCTGACCAAGTTTCTGTCACTGCGCATGGTGTTCGAA	4527
Dd	4716	ACAGATTAATACATTAATTAATCTTATGACAAAGATTCCTTAATGTAAGATATGTTTTGAA	4775
OY	4528	AACCTCGCAAAACAGTGTGAGCAATCTGTCTGA-CGGCAACCTGCAAGGCACTGTGAA	4586

[illegible]

CC provide multicomponent, multistage vaccines due to their expression
of sporozite, liver stage, blood stage and sexual stage proteins.
XX
SO Sequence 5181 BP: 2287 A; 684 C; 707 G; 1503 T; 0 other:

Query Match 25.5%; Score 1262; DB 16; Length 5181;
Best Local Similarity 56.0%; Pred. No. 1.9e-296;
Matches 2910; Conservative 0; Mismatches 1995; Indels 291; Gaps 16;

```
QY 10 ATGAAATCATTTTCTCTCTGTTTCATTTTCTTTTATCATCAATACACAGTCCGTG 69
DB 1 ATGAAGATCATATTTCTTTCTATGTTTCATTTCTTTTATTTATTAACAAATGTGTA 60
QY 70 ACCGAGATCCTATCAGAGAGTGTAAAGAACTGAAGCTTTGAGAGATCCCTCCTT 129
DB 61 ACACATGAAAGTATACAAAGACTTCAAAAACAGTAAAGCTTTTGAAGATGCAATTTG 120
QY 130 ACCGATACAGCCTGTTCCAGAGAGAGATGCTGAATGA----- 174
DB 121 ACAGTATGTTTATTTATCATAGGAAAAATGATCTTAATGAAGAAATTAATCTACA 180
QY 175 -----GGGACAGTGGCAGGCGCTTACACACAGCACACCGGT 213
DB 181 AAAGTGCAATGCTCAAGTGTGTACAGTGTGTACAGTGTGTACAGTGTGTGTGTGT 240
QY 214 TCTAAAGGCTGTGGCTAGCGGTGGCTCCGGTGGCTGTGGCTGTGGGGTTCGGTC 273
DB 241 ACAAGTGTACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
QY 274 GCTTCGGGCGCAGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 333
DB 301 GGTACAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
QY 334 CGAAGAACCAATCCATGTGACA----- 355
DB 361 AGTCATATCTCTGTTCAACACTTACCTGTTCAAAATCTTCATCTGTGTGTGTGTGT 420
QY 356 -----ACTTAGGATTCGAGCGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 408
DB 421 CCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
QY 409 AACTATCTCTCTATATGAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 468
DB 481 AATTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
QY 469 ATGCGACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 528
DB 541 ATGTTAACTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
QY 529 ATCAATGAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 588
DB 601 ATTATGATATATATATATATATATATATATATATATATATATATATATATATATAT 660
QY 589 GACGTTTGGCCCAATGACTATTTGTCAATTCATTTGAAGATGAGAGCCCAAGAG 648
DB 661 GATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
QY 649 TTGAGACGATTTGAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 708
DB 721 TTAGACGATCTTAAAAAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
QY 709 AATGAGGAGAGATGAGATTTATTTAAAGAAATGAAGAGAGAGAGAGAGAGAGAGAG 768
DB 781 AATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 769 GAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 828
DB 841 GAATTAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 829 GAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 888
DB 901 GAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
```

```
QY 889 GAAGCCATTAACCTCATCAGCGTACTGGAGAGAGCCATGACACCCCTCAAGAGAGAGAG 948
DB 961 GAGACATTAATTTATTAAGGCTTTTACAAAAACGATATGACACTTTAAAAAAAATGAA 1020
QY 949 AATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1008
DB 1021 AACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1009 GGGAGACCCCTTACACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1068
DB 1081 GGAATTAACCAAAATACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
QY 1069 ATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128
DB 1141 ATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1129 GAGCTGAGTACTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1187
DB 1201 GAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1188 -----GACAAAGAGATCAACCAACCTATGAATATCCCAATGTGTGTGTGTGTGT 1239
DB 1261 GATCTACGAAATCTGTTCAAATACCAAAAGTTCCTTATCCAAATGTGTGTGTGTGTGT 1320
QY 1240 CTGTCTTATTAAGATATCAACA-----CGCTCTCAACGAGCTCAATAGCTTCGT 1290
DB 1321 TTACACATCAGATATATCATATATCATATAGCTGTGTGTGTGTGTGTGTGTGTGTGT 1380
QY 1291 GACTTGAATTAACCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1350
DB 1381 GATTTAAATGAATCCGATATCAAAAGAAAAATTAATGAAGAAAAATTAATGAGATTAAG 1440
QY 1351 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1410
DB 1441 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1411 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1470
DB 1501 AATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1471 TATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1530
DB 1561 TATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1531 AACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1590
DB 1621 GTCTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
QY 1591 AATACCTTGGATCTATGAGAGATTTCTAAGCATATATTTGAAGAGAGAGAGAGAGAG 1650
DB 1681 AATAATTTTTCATCTCTATATATATATATATATATATATATATATATATATATATATAT 1740
QY 1651 AAGTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1710
DB 1741 TCATATCTTGAAGAGATTTCTTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1711 AAGATCTCATATGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1770
DB 1801 TATATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1771 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1830
DB 1861 AGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1831 ATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1890
DB 1918 TCCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
QY 1891 ATTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1950
DB 1978 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2037
QY 1951 CATGTGCCGAGATTTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2010
```



```

QY 4054 CCATACAGATTCTCCATTAAGAGAGAGGATAAATTTCTGTCTGTATACACTATATC 4113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4269 TCATTAAATATATGATATTCAGACACAAAAAACACACTTTTAAAAAGTTACAAATATTA 4308
QY 4114 AAGCACTCCATCGACACCGATATCAATTTGCTTAATGATGTGCTGGGTATTTACAAGATC 4173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4309 AAAGAAATCAGTAGAAAAATGATTAATTTGACAGAGAGATTAATGATTAATGAAAAAG 4368
QY 4174 CTGAGCGGAAAAATCAACTGTGACCTTGACTATTAATAAGTATATAC----- 4224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4369 GTTTTAGGAAATATATAGGATGATTTAGAAATCAATTTAAAAAGTTATCAAGAGAAAAAG 4428
QY 4225 -----GATPAGCAA 4233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4429 GAGAAGTTCCATCATCATCACCAACAACACCTCCGTACACAGTAAAAACAGACACACAA 4488
QY 4234 GCGGAAATGAAAAATATCTGCCCTCTCTGAATTAACATCGAAACCTGTACAGACAGTGTG 4293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4489 AAGAAGGAAAGTAAAGTCTTCCATTTTAAACAAACATTTGAGACCTTATACAAATACTTA 4548
QY 4294 AACGCAAAATCGACCTTGTGTATTCACCTGGAGGCCAAGGTCTCACTATATCTTAC 4353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4549 GTTATATAAATGAGGATTAATTAATTAAGGCAAGATTAACGATTTGATTAATCTT 4608
QY 4354 GAGAAGACCAATGTGGAAGTTAAATCAAGAGCTGAACCTCACTCAAAACATCCAAAGAC 4413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4609 GAAAAAGATGAGACACATGTTAAAAATTAATACTTAATTAATTAAGCAATTTGATATAC 4668
QY 4414 AAGCTGCAGATTTCAAGAAAAATTAACAAATTCGTGCAATTTGACAGCTGTCTACCGAT 4473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4669 AAATATAGATCTTTTAAAAACATPAAGACTTCGAAAGCAATTTAAAAATTTGATTAATGAT 4728
QY 4474 TATAACCAACAACTCTCTGACCAAGTTTCTGTCCACTGGCATGTGTGTAACAACTC 4533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4729 GATACGAAAAAGATATCTGGCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 4785
QY 4534 GCCAAAACAGTGTGAGCAATCTGTGACGCAACCTGACGGGCACTGTAACATCTCC 4593
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4786 CCTAATACAAATATATATCAAAATTAATTAAGAAAAATTCGAAGATATGTTAAACATTTCA 4845
QY 4594 CAGCAACCAATCGTGAAGAAACAGTCCCGCAGAAATAGCGGCTGTTCAGGCATCTGGAC 4653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4846 CAACACCAATCGTAAAAAAACAAATGTCAGAAAAATTCGTGATGTTTCAGACATTTAGAT 4905
QY 4654 GAGGCGCAAGAGTGCAGATGTCTCTGACACTACAAACAGAAAGAGATTAAGTGCCTGGAG 4713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4906 GAAAGAGAGAAATGTAAATGTTTATTAATTAACAAACAGAAAGGATTAATGTTGAA 4965
QY 4714 AACCAAAACCTTACCTGCAATGAAGAAACATGCGGGGTGACGCCGATGCTAAATGACAC 4773
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4966 AATCCAAATCTTACTTGTAGACGAAATTAATGCTGTGATGTGATGCAAGATGCAAAATGTAAC 5025
QY 4774 GAGGAAGACAGCGCTTACAGGAAAGAAATCAACATGCGAGTGTACTTAAGCCGCACTCC 4833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5026 GAAGAAAGATTCAGTAGCAAGCAAGAAAGAAATCAATGTGATGACTAAACCTGATTTCT 5085
QY 4834 TATCCACTCTTCAGAGGATTTTTCCTCCAGCTCTAATTTCTGCGCATCTCCTCTGTG 4893
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5086 TATCCACTCTTCAGAGGATTTTTCCTCCAGCTCTAATTTCTGAGAAATATCAATTTCTTA 5145
QY 4894 CTGATCTCTGATGATGATCTGTACAGCTTCACTTAA 4929
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5146 TTAATACATGTTAATATATATACAGTTTCAATTAA 5181

```

RESULT 5
 AAT97956 standard; DNA; 1950 BP.
 ID AAT97956;
 AC AAT97956;
 XX
 DT 03-APR-1998 (first entry)

DT 03-APR-1998 (first entry)

```

XX XX Chimeric MSA-1 antigenic protein 1 used in a malaria vaccine.
DE DE Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
XX XX MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
KW KW Immune response; humoral; cell-mediated; merozoite; ss.
XX XX Chimeric - Mammalia.
OS OS Chimeric - Plasmodium falciparum.
OS OS Chimeric - Homo sapiens.
XX XX
FH FH Key
FT FT Location/Qualifiers
FT FT 1..165
FT FT /tag= a
FT FT /note= "signal sequence added to improve
FT FT immunogenicity"
FT FT misc_feature
FT FT 166..1893
FT FT /tag= b
FT FT /note= "Plasmodium falciparum MSA-1 peptide"
FT FT 1894..1950
FT FT /tag= c
FT FT /note= "anchor sequence added to improve
FT FT immunogenicity"
PN PN W09726911-A1.
XX XX 31-JUL-1997.
XX XX 29-JAN-1997; 97WD-US01395.
XX XX 29-JAN-1996; 96US-0593006.
XX XX (GEOU) UNIV GEORGETOWN.
XX XX Davidson EA, Yang S;
XX XX WPI: 1997-393372/36.
XX XX Malaria vaccine - comprises expression vector expressing fragment of
PT PT merozoite surface antigen
XX XX
PS PS Claim 33; Fig 2; 75bp; English.
XX XX
CC CC The present sequence encodes a chimeric protein that contains a
CC CC mammalian signal and anchor sequence, and a merozoite surface
CC CC antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of
CC CC MSA-1 provides a more specific response than the complete MSA-1,
CC CC and attachment of anchor and signal sequences improve the immunogenicity
CC CC of the protein better than the use of an adjuvant. The chimeric protein,
CC CC and expression vectors (analogues that express MSA-1 without either
CC CC signal or anchor peptides), particularly in the form of recombinant
CC CC vaccinia virus, are used in vaccines to prevent or treat malaria caused
CC CC by Plasmodium falciparum. The vaccinia vector expresses the antigen
CC CC fragment for many days, or even years, generating a long-lasting immune
CC CC response (humoral and/or cell-mediated) against the merozoite form of
CC CC the parasite, in humans or other animals.
XX XX
SO SO Sequence 1950 BP: 830 A; 236 C; 277 G; 607 T; 0 other:
Query Match 20.2%; Score 1000; DB 18; Length 1950;
Best Local Similarity 72.4%; Pred. No. 6.8e-233;
Matches 1297; Conservative 0; Mismatches 495; Indels 0; Gaps 0;
QY 3138 AAGCAACTCAGTCACTGAAACCAATCCGAAACAGTACTGCAAGAACTTCTAGTGTCTT 3197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 AAATGAATGTAATCACTTAATTAACCCAAACAGCATATTAACAAACCTTCTGTTTCTT 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3198 CAACAAGAGAAGAGAGCGGATCGCGAGACAGAGAACTCTGGAGAACACCAAGAT 3257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 TACACAAAAAAGAGAGCTGAATAGCAGAACTGAAAACACATTTAGAAAAACAAAAAT 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3258 TCTTCTCAACACTACAAAGCGCTGTCAGTATTAATTAATGCGAGAGTCTCTCGAA 3317
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 279 ATTATTGAACATTAATAAGACCTGTTAAATATATATATATGATGATCATCTCCATTAA 338
Oy 3318 GACCTCTCCGAGAGAGATCAGACCGAGATTAATCACTGCGACGCTGAGAACTTCAA 3377
Db 339 AACCTTAAGTGAAGATCAATTCACACGAGAGATTAATTTCGCAGTTTGAAAACTTTAA 398
Oy 3378 GGTCTGCTGACGTGAGGAGGACGAGACAACTGAACCTGAGAGAGAGAGAGT 3437
Db 399 ACTATTAAATTAAGTGAAGAAAAATTAAGATTAATTAAATTTGAAAAAGAAAAATTT 458
Oy 3438 CAGCTACCTCTAGCGGACGTGATCCTGATCGCGGAGCTCAGAGAGTCAATTAAGA 3497
Db 459 ATCAATCACTTATCAAGTGGATATCATCTTAATTTGCTGAATTAAGAGATTAATAAAAA 518
Oy 3498 CAGAACTACACCGGACCAAGGCAAGAGATTAATACAGAGTGAATACGACTGGA 3557
Db 519 TAAAAATTTATACAGTAATTTCTCAGAGTAAATATACGATGTTAAACAAATTCATTAA 578
Oy 3558 ATCTTACAAAGAGTCTCTGCTGAGAGAAACAGATGTGCGCACTGTGTCTGAATCTGG 3617
Db 579 ATCTTACAAAGAGTCTCTGCGAAGAGACAGATGTTGCAACAGTGTGAAGTGAAGTGG 638
Oy 3618 CTCGACACACGAGGACGATCTCAACCTAGAAAGCCTGCATCTACTCATGTGAGACCGA 3677
Db 639 ATCCGACACATTAGAACAAAGTCAACCAAGAAACCGACATCACTCATGTAGGACAGA 698
Oy 3678 GTCCATACATTAATACACATCTCAGAACCTGACGATGAGTGCATGACGTCATCTGT 3737
Db 699 GCTTAACACATTAACACAAATGCAAAATGTCGATGATGAAGTATGATGATATCATGT 758
Oy 3738 GCTTATCTTCCGCGAGAGGAGGAGGACATGACATGACCTGCGGACAGTGTGACCGGTA 3797
Db 759 ACCTTATTTGGAGATTCGAGAAAGATTAATGATGATTTAGCAAGTGTGAACAGAGAA 818
Oy 3798 GCGTGCATCTCTCCGATGATGATTAATCTGTCACAAATGAGAGAAATACGAGT 3857
Db 819 ACAGTAACTCTCTCCGATGATGATTAATCTGTCACAAATGAGATATAGT 878
Oy 3858 GCTTATCTGAACCTCTGCGAGGCGTATAGTCTCTCAAGAAACAGCTGAGAGATA 3917
Db 879 TTTTATTTTAAACCTTTAGCAGGTGTTTATAGAGTTTAAAAACATTTAGAAAAATTA 938
Oy 3918 CCGTATGACCTCAATGTCAGGTGAGGACATTTCTGAACACGCGCTTTAATTAAGAGA 3977
Db 939 CCGTATGACCTTAAATGTTAATGTTAAGATATTTTAAATTCACGATTTAATAAACGTA 998
Oy 3978 AAATTTCAAGACGCTTGAGAGGACCTGATTCCTCTAAAGACCTGACCTCCCTCAA 4037
Db 999 AAATTTCAAAAGTGTTTAGAAATCAGATTTAATTCATATAAGATTTAACATCAAGTAA 1058
Oy 4038 CTACGTTGTCAAGAGCCATACAGTTCTCTCAATAAAGAAAGGAGTAAATTTCTGTCT 4097
Db 1059 TTATGTGTCAAGAGATCCATATAATTTCTTATATAAGAAAAAGAGATATTTCTTAAG 1118
Oy 4098 TATGTTACAACTATATCAGAGACTCCATGACACCATATTCATTTGCTAATGATGTGCT 4157
Db 1119 CAGTTATATATATATTAATGAATTCATATAGATAGATGAATTAATTTTGAAGAAATGTCT 1178
Oy 4158 GGGGATTTCAAGACCTGAGGAAAAATACAAAGTCTGCTTGAATTAATAAAGTA 4217
Db 1179 TGGATATTTAAATAATTTATCCGAAAAATATATAAGATTTAGATTCATTAATAAATA 1238
Oy 4218 TATCAACGATTAAGACGAGGAGATGAATAATATGCTCCCTCTGTAATACATGCAAGAC 4277
Db 1239 TATCAACGCAAAACAGGAGGAAAAATGAGAAATACCTTCCTTTTAAACAAATTTAGAC 1298
Oy 4278 CCGTACAAAGAGTGAACGACAAATGACCTTTGTAATTCACCTGAGGCGCAAGT 4337
Db 1299 CTTATATATAACAGTTAATGATTAATTAATTTGTAATTCATTTGAAGCAAAAGT 1358
Oy 4338 CCTCAACTTACTTACGAGAGAGACATGTGAAGTTAAATCAAGAGACTGACATCT 4397
Db 1359 TCTTAATTTATCATATATGAGAAATCAAAAGTGAAGTTAAATAAAGAACTTAATTACTT 1418

```

```

Oy 4398 CAAACAAATCCAAGACAAACCTGGCAGATTTCAAGAAAAATTAACAATTTCTGGAATTC 4457
Db 1419 AAAACAAATCCAAGACAAATTTGGCAGATTTTAAAAAAATTAACAATTTCTGGAATTC 1478
Oy 4458 AGACCTGTCTACCGATTAATACCAACAATCTCTGACCAAGTTTCTCCACTGGCAT 4517
Db 1479 TGAATTAATCAACAGATTAATTAACCATTAATTAATTAATTAATTAATTAATTAATTA 1538
Oy 4518 GGGTTGCAAAACCTGCGCAAAAGTGTGAGCATGTGCTGACGCGGACCTGAGG 4577
Db 1539 GGTTTTGAATTCCTGCTTAACACGTTTATATCTAATTTACTTGATGAAACTTGCAATG 1598
Oy 4578 CATGCTGAACATCTCCACACCAATGCTGAGAAAGAGTCCGCCAGATTAAGCGGCTG 4637
Db 1599 TATGTTAAACATTTTACACACCAATGCGTAAAAAAACAATGTCACAAATTTCTGATG 1658
Oy 4638 TTTTCAGGCACTCTGAGACGCGGAGAGTGCAGAGTGTCTCTGAACTACAAACAGAG 4697
Db 1659 TTTTCAGCACTTTAGAGAAAGAGAAATGTAATGTTTATTAATTAACAAACAGAGAG 1718
Oy 4698 AGATTAAGTCTGAGAGAACCAACCCCTACCTGCAATGAACAAATGGCGGTGACGC 4757
Db 1719 TGATTAATGTGTGAATAATCCAATCTCTTGTAAAGAAATTAATGATGATGATG 1778
Oy 4758 CGATGCTAATSCACCGAGAGACAGCGCTCTAAGCAAGAAAGAAATCAATCGAGTG 4817
Db 1779 AGATGCAATGTACCGAAGAGATTCAGGTAGCAACGAGAAAGAAATCAATGATG 1838
Oy 4818 TACTTAAGCCCGCTCTATCCACTGCTGACGAGGATTTTCTCAGCTCAATTTCTCT 4877
Db 1839 TACTTAACCTGATCTTATACACTTTTTCATGATGATTTTCTGAGTCTCTCACTTCTT 1898
Oy 4878 GGGCATCTCTCTGCTGATCTCATGCTGATCTGATCTGATCTGATCTGATCTGATCT 4929
Db 1899 AGGAATATCTCTTATTAATCTACTGATTAATTAATTAATTAATTAATTAATTAATTA 1950

RESULT 6
ID AAT97958
AAT97958 standard; DNA; 1897 BP.
XX
AC AAT97958;
XX
DT 03-APR-1998 (first entry)
XX
DE Chimeric MSA-1 antigenic protein 3 used in a malaria vaccine.
XX
KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
KW MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
KW Immune response; humoral; cell-mediated; merozoite; ss.
OS Chimeric - Plasmodium falciparum.
OS Chimeric - Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1..1839
FT FT /tag= "a"
FT FT /note= "Plasmodium falciparum MSA-1 peptide"
FT misc_feature 1840..1897
FT FT /tag= "b"
FT FT /note= "anchor sequence added to improve immunogenicity"
XX
PD 31-JUL-1997.
XX
PD 29-JAN-1997; 97WO-US01395.
XX
PR 29-JAN-1996; 96US-0593006.
XX
PA (GEOU ) UNIV GEORGETOWN;

```



```
OY 4278 CCTGTACAGACGATGACGACCAATTCAGCCCTTCTGTAATTCACCTGGAGGCCAAGCT 4337
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1299 CTTATATTAACAGATTATGATTAATTAATTTATTTGTAATTCATTGGAACCAAAAGT 1358
OY 4338 CCTCACTTACTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4397
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1359 TCTAATATATATGATGAAATTCAAACGTAGAGTTAAATTAAGAACTTAATTAATGCT 1418
OY 4398 CAACAATTCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4457
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1419 AAAAACAATTCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1478
OY 4458 AGACGCTGTACGATTAATACCAATTCCTGACCAAGTTCTGCTCCAGCTGACAT 4517
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1479 TGTATTTATACAGATTTATACCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1538
OY 4518 GGTGTTCGAAACCTGCGCAAAACAGTGTGACCAATCTGCTGACGACGACGACGACGACGAC 4577
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1539 GGTGTTCGAAACCTGCGCAAAACAGTGTGACCAATCTGCTGACGACGACGACGACGACGAC 1598
OY 4578 CATGCTGAACATCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4637
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1599 TATGTTAAATCTTCAACACCAATGCGTAAACCAATGCTCAACAAATTCGTGATG 1658
OY 4638 TTTTACGACATCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4697
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1659 TTTTACGACATCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1718
OY 4698 AGATTAAGTCGTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4757
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1719 TATTAATATGTTGAAATTCCAAAATCTCTGTAACGAAATTAATGATGATGATG 1778
OY 4758 CATGCTGAATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4817
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1779 AGATGCAATATGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1838
OY 4818 TACTAAGCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4873
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1839 TACTAAGCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1894

RESULT 8
AA56008
ID AA56008 standard; cDNA; 1065 BP.
XX
AC AA56008;
XX
DT 31-AUG-1999 (first entry)
XX
DE Mezoizite surface protein MSP-1-42 modified cDNA.
XX
KW MSP-1; mezoizite surface protein; malaria; vaccine;
KM protein engineering; codon usage; transgenic animal; ss.
XX
OS Plasmodium falciparum.
XX
PN MO9920774-f2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
XX
PR 20-OCT-1997; 97US-0062592.
XX
PA (GEN2 ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI: 1999-28813/24.
XX
P-PSDB: AA509372.
```

```
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Claim 1; Fig 1; 35pp; English.
XX
CC This novel, modified nucleic acid encodes the 42 kDa C-terminal
CC portion (see AA509372) of malaria mezoizite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AA56009)
CC such that 306 nucleotide positions have been replaced to lower the
CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability
CC motifs while maintaining the same protein amino acid sequence.
CC These alterations allow MSP-1-42 to be expressed in mammalian cell
CC culture and in transgenic mice. Native MSP-1-12 is known to be
CC difficult to express in cell culture systems, mammalian cell
CC culture systems or in transgenic animals. The invention allows
CC expression of MSP-1-42 protein in the milk of transgenic animals,
CC and also provides a DNA vaccine comprising a vector containing the
CC altered MSP-1-42 sequence.
XX
SQ Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;
XX
Query Match 15.1%; Score 744.2; DB 20; Length 1065;
Best Local Similarity 81.3%; Pred. No. 9, 3e-171;
Matches 863; Conservative 0; Mismatches 198; Indels 0; Gaps 0;
OY 3799 GCTGTACATCTTCCCTGATGATTAATTCATTCGAAATCGAAGCAATGCAAGT 3858
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 GCCGTACGTCCTCTGCTGATGATTAATTCATTCGAAATCGAAGCAATGCAAGT 60
OY 3859 CTCTATCTAAACCTTGGCAGCGCTGTATAGTCTCTCAAGAACAGCTGAGAAATAC 3918
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 CTGTACTGTAAGCGCGTGCGAGGCGTACCGGAGCTGAAGAGAGCGTGAACAC 120
OY 3919 GTGATACCTTCAATGTCACGTAAGGACATTTCTAAGACCGCTTATTAAGAGAA 3978
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 GTGATACCTTCAATGTCACGTAAGGACATTTCTAAGACCGCGTTCACACGCGGAG 180
OY 3979 AATTTCAAGACGCTTGGAGACGACGATGATTCCTTAAGACGCTGACCTCTTAAC 4038
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 181 AACTTCAAGACGCTGCGAGAGCGATGATTCCTTCAAGACGCTGACCTGACACCAAC 240
OY 4039 TACGTGTCAAGAGCCATACAGATTCCTCAATTAAGAGAGAGGATTAATTTCTGTCT 4098
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 241 TACGTGTCAAGAGATTCCTCAATTAAGAGAGAGGATTAATTTCTGTGTCTGAGC 300
OY 4099 AGTTACACTATATCAAGGATCCTACGACACCGATATCAATTTCCCTAATGATGCTG 4158
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 AGTTACACTATATCAAGGATCCTACGACACCGATATCAATTTCCCTAATGATGCTG 360
OY 4159 GGGTATTAAGATCCTGAGCGGAAATTAAGTGTGACCTGACCTGATTAATAAGTAT 4218
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 361 GGTATTAAGATCCTGAGCGGAAATTAAGTGTGACCTGACCTGATTAATAAGTAT 420
OY 4219 ATCAACGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4278
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 421 ATCAACGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
OY 4279 CTGTACAGACGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4338
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 481 CTGTACAGACGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 540
OY 4339 CTCACATATATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4398
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 541 CTGACATTAACATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
OY 4399 AAAACAATCAAGACGATGCGAGATTCAGAAATTAAGCAATTTCTGGAATTTGCA 4458
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 601 AAGACCATCAAGATTAAGTGGCGGATTCAGAAAGAACACAACTTCGCGGATCGCC 660
OY 4459 GACCTGTCTACGATTAATACCAACAACTTCCTGACCAAGTTCTGTGACATGCGCATG 4518
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 661 GATCTGAGCAGCAGATTAACCAACCAACCACTGTCACCAAGTTCTGTGACACCGGATG 720
```



```

Db 781 ATGCTGAACATCAGCCAGCCAGCAGTGTGTGAAGAAGACAGTGTCCCGAGAACAGCGGGTGT 840
Oy 4639 TTCAGGCATCTGGACGCGGGAAGATGCAAGTCTCTCTGAACTACAAACAAGAGA 4698
Db 841 TTCAGACCTGTGATGAGAGAGAGTGTAACTGTCTGCTGAACATACAGCAGAGGT 900
Oy 4699 GATAAGTCTGTGAGAACCCCAACCCCTACCTGCAATGAAGAACATGCGGTGTAGGCC 4758
Db 901 GATAGTGTGTGGAAGAACCCCAATCTACTTGTAAAGAACAAATGTGTGTATGCC 960
Oy 4739 GATGCTAAATGCACCGAGAGACAGCGGCTCTTAACGGAAGAAATACATGCGAGTGT 4818
Db 961 GATGCAAGTGTACCGAGAGAGATTCAGGAGACCAAGGGAAGAAATACATGCTGAGTGT 1020
Oy 4819 ACTAGCCCGACTCCTATCCACTCTTCGACGCGGATTTTGTG 4859
Db 1021 ACCAAGCCTGATCTTATTCACCTGTGATGTATCTTCTGTG 1061

RESULT 10
AA56021
ID AA56021 standard; cDNA: 1140 BP.
AC AA56021:
XX 31-AUG-1999 (first entry)
DT 31-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42 modified cDNA.
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KM transgenic animal; mutant; ss.
XX Plasmodium falciparum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1131
FT /tag= a
FT sig_peptide 1..45
FT /tag= b
FT mat_peptide 46..1128
FT /tag= c
XX
PN MO9920774-A2.
XX
XX 29-APR-1999.
PF 20-OCT-1998; 98WO-US22226.
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ ) GENZYME TRANSGENICS CORP.
PI Chen LH, Meade H;
XX WPI: 1999-288313/24.
DR P-PSDB; AAY09374.
XX
XX Modified malarial protein for use in anti-malarial vaccines
XX
XX Example; Fig 11; 35pp; English.
XX
XX This novel, modified nucleic acid encodes a 42 kDa C-terminal
XX pat (see AAY09374) of malaria merozoite surface protein MSP-1
XX (MSP-1-42), an important target for the development of a vaccine
XX against Plasmodium falciparum. The nucleic acid sequence has been
XX modified compared to the native sequence of MSP-1-42 (see AA56009)
XX such that the AT content has been reduced and 10 mRNA instability
XX motifs eliminated while maintaining the protein amino acid sequence.
XX In addition, a sequence encoding a 15-amino acid beta-casein signal
XX peptide has been added to the 5' end of the sequence, and N262Q

```

```

CC and N181Q mutations have been introduced to eliminate
CC N-glycosylation sites. These alterations allow MSP-1-42 to be
CC expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC invention also provides a DNA vaccine comprising a vector containing
CC an altered MSP-1-42 sequence.
XX
SQ Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;
Query Match 15.0%; Score 742; DB 20; Length 1140;
Best Local Similarity 80.8%; Pred. No. 3.3e-170;
Matches 865; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
Oy 3791 CCGGTGAGCGCTGCTCACTCCTCCGTGATGATTAATCTGTGCCAAATCGAAGCAAT 3850
Db 38 CCATTGCAAGCGCTCACTCCTCCTGATGATTAATCTGTGCCAAATCGAAGCAAT 97
Oy 3851 ACGAAGTGTCTATCTGAAACCTGTGGCAGCGGTCTATAGTCTCTCAAGAAACGCTGG 3910
Db 98 ACGAGTGTCTATCTGAAACCTGTGGCAGCGGTCTATAGTCTCTCAAGAAACGCTGG 157
Oy 3911 AGAATACGCTGATGACCTTCAATGTCACAGCTGAAGACATTTGTGAACGCCCTTAAAT 3970
Db 158 AGAACAAGCTGATGACCTTCAAGAAACGCTGAAGAGATATCTGTGAACGACAGGTTCAACA 217
Oy 3971 AGAGAGAAATTTCAAGAACTCTTGGAGAGCGACTTGTCTCTATAAGACCTGACT 4030
Db 218 AGAGGAGAACTTCAAGAAACGCTGAAGAGCGATCTGATCCCTACAAAGATCTGACCA 277
Oy 4031 CCTTAATCACTGTTTCAAGAACCCATACAGTTCTCTCAATTAAGAAAGAGATTAAT 4090
Db 278 GCAGCAACTACCTGTTCAAGAAATCCCTCAAGTTCTGTGAACGAGAAAGAGATTAAGT 337
Oy 4091 TTCTGTAGTTACAACTATATCAAGAGACTCCATGACACGATATTCATTTCCGTAATG 4150
Db 338 TCTGAGCACTTACAACTATATCAAGAGATGATGATGATGACACGATATTCATTTCCGTAACG 397
Oy 4151 ATGCTCTGGGATTTACAAAGATCTGTAGGAAATATACAGTCTGACTTGAATATTA 4210
Db 398 ATGCTCTGGGATTTACAAAGATCTGTAGGAAATATACAGTCTGACTTGAATATTA 457
Oy 4211 AAAAGTATATCAAGATTAAGCAGGCGAGATGAATAAATATATGCTTCTGTAATACA 4270
Db 458 AGAAGTACATCAACGATTAAGCAGGAGAAACGAAAGATCTGCCCTTCCGAACAACA 517
Oy 4271 TCGAACCCTGTACAAAGCAGGACGAAACAAATGACCTCTGTAATTCACCTGGAGG 4330
Db 518 TCGAACCCTGTACAAAGCAGGACGAAACAAATGATGATGATGATGATGATGATGATG 577
Oy 4331 CCAAGTCTCTCACTATCTTACGAGAGAGCAATGTGGAATTAATAATCAAGAGCTGA 4390
Db 578 CCAAGTCTCTGATACATATGAGAGAGCAAGCTGAGGTCAAGATCAAGAGAGCTGA 637
Oy 4391 ACTACCTCAAAACAATCCAAAGCAAGCTGGCAGATTTCAAGAAATTAACAATTTCTGC 4450
Db 638 ATTACCTGAAGACCATCAGATTAAGCTGGCAGATTTCAAGAAATTAACAATTTCTGC 697
Oy 4451 GAATGCAAGACTGTCTACCGGATTTAACCACCAACCAATCTCTGACCAAGTTCTGTGCA 4510
Db 698 GAATGCAAGACTGTGACCGGATTTAACCACCAACCAACCTGTTGACCAAGTTCTGTGCA 757
Oy 4511 CTGGCATGGTGTTCGAAAACCTCGCCAAACAGTGTCTGACCAATGTGCTGAGGCAAC 4570
Db 758 CCGGAATGGTGTTCGAAAACCTCGCCAAACAGTGTCTGACCAATGTGCTGAGGCAAC 817
Oy 4571 TCGAGGCAATGCTGAACATCTCCAGCACCAATGCTGGAAGAAACATGCCCAAGATTA 4630
Db 818 TCGAGGCAATGCTGCAAGATTCACCGACCAAGTGTGTAAGAAAGAGTGTCCCGCAACA 877
Oy 4631 GGGGCTGTTTCAGGATGTGAGCAGCGCGAAGAGTGAAGTGTCTGTAATCTCAAAC 4690
Db 878 GGGGCTGTTTCAGGATGTGAGCAGCGCGAAGAGTGAAGTGTCTGTAATCTCAAAC 937
Oy 4691 AAGAAGAGATTAAGTGTGAGAACCCAAACCTTACTGTCAATGAATAAATATGCGGGT 4750

```

```

Db 938 AGGAAGAGATAGTGTGTGAAACCCCAATCTTCTAAGAGAACATGAGAGAT 997
Oy 4751 GTGACCCCATGCTAATATGACCGAGAGACGCGCTTAACGAAAGAAATACAT 4810
Db 998 GCGATCCCGATGCGCAAGTGTACCGAGAGATTCAGAGAACGAAAGATACCTT 1057
Oy 4811 GCGAGTGTACTAAGCCGACCTCTATCAGCTCTTCGACGAGATTTTTCG 4860
Db 1058 GCGAGTGTACCAAGCCTGATTTCTATCAGCTCTTCGATGATTTTCTG 1107

RESULT 11
AAK25593
ID AAK25593 standard; cDNA; 1140 BP.
XX
AC AAK25593;
Dt 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42 modified cDNA.
XX
KM MSP-1; merozoite surface protein; malaria; vaccine;
KM protein engineering; protein expression; codon usage;
KM Transgenic animal; mutant; ss.
OS Plasmodium falciparum.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1131
FT sig_peptide 1..45
FT mat_peptide 46..1128
FT /*tag- b
/*tag- c
XX
PN WO9920766-42.
XX
PD 29-APR-1999.
XX
PE 20-OCT-1998; 98WO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GEN2 ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI: 1999-302742/25.
DR P-PSDB: AAY05834.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
PS Claim 12: Fig 11; 43pp; English.
XX
CC This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC part (see AAY05834) of malaria merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AAK25587)
CC such that the AT content has been reduced and 10 mRNA instability
CC motifs eliminated while maintaining the protein amino acid sequence.
CC In addition, a sequence encoding a 15-amino acid beta-casein signal
CC peptide has been added to the 5' end of the sequence, and N262Q
CC and N181Q mutations have been introduced to eliminate
CC N-glycosylation sites. These alterations allow MSP-1-42 to be
CC expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC invention provides modified recombinant nucleic acid sequences and
CC methods for increasing the mRNA levels and protein expression of
CC proteins that are difficult to express in cell culture systems,

```

```

CC mammalian cell culture systems or in transgenic animals. The
CC preferred difficult protein candidates for expression are those
CC derived from lower organisms such as parasites, bacteria and
CC viruses that have DNA coding sequences of high AT content or which
CC have mRNA instability motifs or rare codons relative to the
CC recombinant expression system to be used. The invention allows
CC expression of MSP-1 in the milk of transgenic animals, and also
CC provides a DNA vaccine comprising a vector containing the altered
CC MSP-1-42 sequence.
XX
SQ Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;
Query Match 15.0%; Score 742; DB 20; Length 1140;
Best Local Similarity 80.8%; Pred. No. 3.3e-170;
Matches 865; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
Oy 3791 CCGGTAGGCTGTCACCTCCGCTGATGATTAATCTGTCCTCAAAATGAGAGAAAT 3850
Db 38 CCAATGCAAGCGTCACTCCCTCGTATGATACATCCGTCAAGATGAGAGAGAGT 97
Oy 3851 ACAGAGTGTCTATCTGAAACCTCTGAGCGCTCTATAGTCTCTCAAGAAACAGCTGG 3910
Db 98 ACAGAGTGTCTATCTGAAACCTCTGAGCGCTCTATAGTCTCTCAAGAAACAGCTGG 157
Oy 3911 AGAATTAAGTATGATACCTTAAATGTCAGCTGAAGACATTTGAAACAGCCCTTTAAT 3970
Db 158 AGAACACAGTGAATGACCTTCAACGTGAAGAGATATCTTGAACAGCAGGTTCAACA 217
Oy 3971 AGAGAGAAATTTCAAGAACGCTTGAGAGAGACATTTGCTTAAAGACCTGACCT 4030
Db 218 AGAGGAGAACTTCAAGAACGCTTGAGAGAGACATTTGCTTAAAGACCTGACCT 277
Oy 4031 CCTCTAAGTATGTTGTCAGAGACCATACAAAGTTCTCTCAATTAAGAGAGGATTAAT 4090
Db 278 GCAGCAACTGCTGTCAGAAAGATCCCTACAAAGTTCTCTGAAACAGAGAGAGATTAAGT 337
Oy 4091 TTCTGTCTAGTTCAACATATATCAAGAGACTCCATGACACCGATATCAATTTGCTAATG 4150
Db 338 TCCTGAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 397
Oy 4151 ATGTGCTGGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4210
Db 398 ATGTGCTGGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 457
Oy 4211 AAAAGTATATCAATGATTAACAGAGGAGATGAAATTAATTAATTAATTAATTAATTA 4270
Db 458 AGAAGTATATCAATGATTAACAGAGGAGATGAAATTAATTAATTAATTAATTAATTA 517
Oy 4271 TCGAAACCTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4330
Db 518 TCGAGACCTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 577
Oy 4331 CCAAGGTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4390
Db 578 CCAAGGTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 637
Oy 4391 ACTACCTCAAAACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4450
Db 638 ATTACCTGAAGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697
Oy 4451 GAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4510
Db 698 GAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
Oy 4511 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4570
Db 758 CCGGATGCTCTTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 817
Oy 4571 TGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4630
Db 818 TGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 877
Oy 4631 GCGGCTGTTTCAGGAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4690

```



```
AC AAX56009;
XX
XX 31-AUG-1999 (first entry)
XX
XX Merozoite surface protein MSP-1-42 cDNA.
DE
XX MSP-1; merozoite surface protein; malaria; vaccine;
KM protein engineering; protein expression; codon usage;
XX transgenic animal; ss.
XX
XX Plasmodium falciparum.
OS
XX Key Location/Qualifiers
FH CDS 1..1086
FT
FT CDS /*tag= a
XX
XX W0920774-A2.
XX
XX 29-APR-1999.
XX
XX 20-OCT-1998; 98MO-US22226.
XX
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX
XX (GEN2 ) GENZYME TRANSGENICS CORP.
XX
XX Chen LH, Meade H;
XX
XX WPI: 1999-28833/24.
XX P-PSDB; AAY09373.
XX
XX Modified malarial protein for use in anti-malarial vaccines
XX
XX Example; Fig 2; 35pp; English.
XX
XX This nucleic acid encodes a 42 kDa C-terminal portion (see AAY09373)
XX of malarial merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The 3' end of the sequence has been
XX modified to include a 6xHis tag. The nucleic acid sequence has
XX been modified (see AAX56008) according to a method of the invention
XX in order to improve expression in mammalian host cells and in
XX transgenic animals. In the modified coding sequence, 306 nucleotide
XX positions have been replaced to lower the AT content (from 76 to
XX 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
XX amino acid sequence is unaltered. In another modified sequence
XX (see AAX56021), a signal peptide sequence has been added and two
XX N-glycosylation sites eliminated. The invention allows expression
XX of MSP-1-42 protein in the milk of transgenic animals, and also
XX provides a DNA vaccine comprising a vector containing the altered
XX MSP-1-42 sequence.
XX
XX Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;
SO
Query Match 12.7%; Score 625.2; DB 20; Length 1088;
Best Local Similarity 74.3%; Pred. No. 8.7e-142;
Matches 789; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
```

```
DB 181 AATTTCAAAAATGTTTATGATCAGATTTATCCATATTAAGATTTACATCAAGTAAT 240
QY 4039 TACGTTGTCAAGGACCCTACCAAGTTCCATTAAGAGAGAGGATTAATTTCTGCT 4098
DB 241 TATGTTTCAAGATTCATTTAAATTTCTTAATTAAGAAAAAGATTAATTTCTTAAGC 300
QY 4099 AGTTACACATATATCAAGAGCTCCATGACACCGATATCAATTTCCGTAATGATGCTG 4158
DB 301 AGTATATATTAATTAAGATTCATTAAGATGATGATTAATTTTGAATGATGCTCT 360
QY 4159 GGGATTTACAAAGATCCGACGAGAAAAATACAGCTGACCTGACCTATTAAGAAAT 4218
DB 361 GGATATTTATTAATATTAATTCGGAATAATTAATCAATTAATTAATTAATTAATTAAT 420
QY 4219 ATCAAGATTAAGCAAGGCGAGATGAATAATATATGCTTCCGTAATTAATCAATCAAGC 4278
DB 421 ATCAAGCAACAACAGGTGAATAATGCAATTAATTAATTAATTAATTAATTAATTAAT 480
QY 4279 CTGTACAGACAGTGAACGACAAATTCGACCTCTGTAATTAATCACTGAGCCAGGTC 4338
DB 481 TTATATTAACACAGTTATGATTAATAATGATTTATTTGTAATTAATTAAGCAAAAAGTT 540
QY 4339 CTGACATATCTTACGAGAGAGCAATGGAAGTTAAATCAAGGAGCTGAATCTACT 4398
DB 541 CTAATATTAATCAATGAGATCAACAGTGAAGTTAAATTAAGAACTTAATTAATCTTA 600
QY 4399 AAAACAATCCAGACAGCTGGCAGATTTCAAGAAAAATTAACATTTCTGCGAATTTGA 4458
DB 601 AAAACAATTCAGCAAAATTTGGCAGATTTTAAAAAATTAACATTTCTGTTGAATTTCT 660
QY 4459 GACCTGTCTACCGATTTATTAACCAACAATCTCTGACCAAGTTTCTGTCCACTGCTATG 4518
DB 661 GATTTATCAACAGATTTTAACCAATTAATTAATTAATTAATTAATTAATTAATTAAT 720
QY 4519 GTGTTCAAAAACCTCGCCAAACAGTGTGAGCAATGCTGCGAACCACTGCAAGGC 4578
DB 721 GTTTTAAAAATCTGTCTAAACCGTTTATCTAATTTACTGATGGAATCTGCAAGGT 780
QY 4579 ATGCTGAACATCTCCAGCAACCAATGCGTAAGAAACAGTGCACCAATTAAGCGCTGT 4638
DB 781 ATGTTAAACATTTTACACAAACCAATGCGTAAGAAACCAATTTCTGCAATGT 840
QY 4639 TTACAGCATCTGACGAGCGGCAAGAGTGAAGTGTCTCTGAATCTAACAAGAAAGA 4698
DB 841 TTACAGCATTTAGATGAAGAAAGAAATGTAATGTTATTAATTAACAAGAAAGGT 900
QY 4699 GATTAAGTGTGAGAGAACCCAAACCTTACCTGCATGAAACCAATGCGGCTGTGACGCC 4758
DB 901 GATTAATGTTGTAATAATCCAAATCTACTTGTATACGAAAAATATATGTTGATGATGA 960
QY 4759 GATGCTTAATGCAACGAGAGAGACGCGCTTAACGGAAGAAATACATGCGAGTGT 4818
DB 961 GATGCCAAATGTACCGAAGAAATTCAGTAGCAACGGAAGAAATCACATGGAATGT 1020
QY 4819 ACTAAGCCGACCTCTATCCACTTTGACGAGGATTTTTCG 4860
DB 1021 ACTAAGCCGATCTTATCCACTTTGATGATGATTTTTCG 1062
```

```
RESULT 14
AAX25587
ID AAX25587 standard; cDNA; 1088 BP.
XX
XX AAX25587;
XX
XX 02-AUG-1999 (first entry)
XX
XX Merozoite surface protein MSP-1-42 cDNA.
DE
XX MSP-1; merozoite surface protein; malaria; vaccine;
KM protein engineering; protein expression; codon usage;
XX transgenic animal; ss.
XX
```


OS	Plasmodium falciparum.
XX	Key
XX	Location/Qualifiers
XX	1..1086
XX	CDS
XX	/*tag= a
XX	W09920766-A2.
XX	
XX	
XX	29-APR-1999.
XX	
XX	20-OCT-1998; 98WO-US22225.
XX	
XX	15-MAY-1998; 98US-0085649.
XX	20-OCT-1997; 97US-0062592.
XX	
XX	(GEN2) GENZYME TRANSGENICS CORP.
XX	
XX	Chen LH, Meade H;
XX	
XX	WPI: 1999-302742/25.
XX	P-PSDB: AAY05833.
XX	
XX	New modified recombinant nucleic acid sequences useful for producing
XX	malaria DNA vaccine
XX	
XX	Disclosure: Fig 2; 43pp; English.
XX	
XX	This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)
XX	of malarial merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX	important target for the development of a vaccine against
XX	Plasmodium falciparum. The 3' end of the sequence has been
XX	modified to include a 6xHis tag. The nucleic acid sequence has
XX	been modified (see AAX25586) according to a method of the invention
XX	in order to improve expression in mammalian host cells and in
XX	transgenic animals. In the modified coding sequence, 306 nucleotide
XX	positions have been replaced to lower the AT content (from 76 to
XX	49.7%) and to eliminate 10 mRNA instability motifs. The encoded
XX	amino acid sequence is unaltered. In another modified sequence
XX	(see AAX25593), a signal peptide sequence has been added and two
XX	N-glycosylation sites eliminated. The invention provides modified
XX	recombinant nucleic acid sequences and methods for increasing the
XX	mRNA levels and protein expression of proteins that are difficult
XX	to express in cell culture systems, mammalian cell culture systems
XX	or in transgenic animals. The preferred difficult protein
XX	candidates for expression are those derived from lower organisms
XX	such as parasites, bacteria and viruses that have DNA coding
XX	sequences of high AT content or which have mRNA instability motifs
XX	or rare codons relative to the recombinant expression system to be
XX	used. The invention allows expression of MSP-1 protein in the milk
XX	of transgenic animals, and also provides a DNA vaccine comprising a
XX	vector containing the altered MSP-1-42 sequence.
XX	
XX	Sequence 1088 BP: 454 A; 139 C; 150 G; 345 T; 0 other:
XX	
XX	Query Match 12.7%; Score 625.2; DB 20; Length 1088;
XX	Best Local Similarity 74.3%; Pred. No. 8.7e-142;
XX	Matches 789; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
XX	
XX	3799 GGTGTCACCTCCCTCCGCTGATTGATTAACATCTGTCGCCAAATGCGAGACGATACGAGTG 3858
XX	1 GCAGTAACCTCCCTCCGTAATGATGATTAACATCTTCTTAATAATTGAAATGAAATACAGAGTT 60
XX	3859 CTCCTATTCGAAACCTCTGCGACGAGCCTATATAGCTCTGTCACAAACAGCTGAGATATAC 3928
XX	61 TTATATTTTAAACCTTTAGCAGAGTCTTTRATGAAAGTTTAAATAAAACATTTAGAAAAATTAAC 120
XX	3919 GTGATGACCTTCATATGTCACACGCTGAGACATTTGCAACAGCCGGTTTAATTAAGACAGAA 3978
XX	121 GTTATGACATTTAATGTTAATGTTAAGATATTTTAAATTCACGATTTAATTAACGTGAA 180
XX	3979 AATTCAAGAACGCTTTGAGACGACGATTTGATTCCTCTATAAAGACCTGACCTCTCTAAC 4038
XX	181 AATTTCAAAAAAGTTTGTGAAATCAATTTAATTAATTAATTAAGATTTAACATCAAGTAAAT 240

OY	4039	TACCTTGTCAAGAGCCCATACATAGTTCCTCAATAAAGAGAAAGAGGATTAATTTCTGCT	4038
Db	241	TATGTGTCAAAAGATCCATTTAAATTTCTTAAATAAAGAAAAAGAGATTAATTTCTTAAGC	300
OY	4099	AGTTACAACATATATACAGAGATCCATCGACACCGATATCAATTTGGCATATGATGTGCTG	4158
Db	301	AGTTATATATATATATTAAGGATTTCAATATGATACGGATATTAATTTTGGCAATATGATTTCTT	360
OY	4159	GGGTATTTACAAAGATCTCGACCGAAAAATACAAAGTCGACCTTGACTATTAATAAAAATAT	4218
Db	361	GGATATTTATTAATAATATTAATCCGAAAAATATTAATATGATTTAGATTTCAATTTAAAAATAT	420
OY	4219	ATCAACGATTAAGCAAGCCGGAATGAAAAATATCTGCCCTTCTGATATACATGAAACC	4278
Db	421	ATCAACGACAAACAAAGGTGAAATGAGAAATACCTTCCCTTTTAAACAAATATTCAGACC	480
OY	4279	CTGTACAGACAGAGAACGACAAATATTCGACCTCTTGATTAATTCACCTGGAGGCCAAAGTC	4338
Db	481	TTATATATTAACAGTTAATGATFAAAATTTGATTTATTTGTATTTCAATTTAGAACCAAAAGTT	540
OY	4339	CTCAACTATACCTTACGAGAGACCAATGTGGAATTTAAATCAAGAGACTGAATCACTTC	4398
Db	541	CTAATATATACATATATGAGAAATCAACAGTAAAGATTAAATTAAGAACTTAATTACTTTA	600
OY	4399	AAAAACAATCCAGACAAAGCTGGCAGATTTTCAGAAAAAATATACAATTTTGTGGGAATTTCA	4458
Db	601	AAAAACAATTCAGACAAATTTGGCAGATTTTAAAAAAAATACAAATTTGTGGGAATTTCT	660
OY	4459	GACCTGTCTACCGATTTATACCAACAATCTCTGACCAAGTTTCTGCTCCACTGGCATG	4518
Db	661	GATTTATACACAGATTTATACCAATTAATTAATTTTACAAAATTTCTTATATGACAGTATG	720
OY	4519	GTGTTCGAAAAACCTGGCCAAAACAGTGTGAGCAATCTGCTCGACGGAACCTGCAGGGC	4578
Db	721	GTTTTGAATAATCTTGCTATAAACCGTTTATCTAATTTACTTGATGAGAAACTTGCAGAGT	780
OY	4579	ATGCTGAACATCTCCCGACGACCAATGCGTGAAGAAACAGTGGCCCCAGATATAGCGGCTG	4638
Db	781	ATGTTTAACCATTTTACACACACATGCGTATTAACAAAACAAATTTGTCACAAAATTTTGATGT	840
OY	4639	TTTCAGGATCTGACGACGCGCGAAGAGTCAAGTGTCTCTGATACATACAAACAAAGAGA	4698
Db	841	TTTCAGACATTTAGATGAGAAAGAGAAGATGTAAATGTTTATTAATTTACAAACAAAGAGT	900
OY	4699	GATATAGTGGTGGAGAACCCAAACCTTACTGCAATGAAACAAATAGCGGGTGTGACGCC	4758
Db	901	GATTAATGTGTGGAATTCCAAAATCCTACTGTTAAAGAAAAATTAAGGTGATGTATCA	960
OY	4759	GATCTTAATATGACCCGAGAGACACGCGTCATACGAGAAACAAATATACATGACAGTGT	4818
Db	961	GATGCTCAAAATGTACCGAAGAGATTTAGGTAGCAACGGAAGAAATAATCATATGTATATGT	1020
OY	4819	ACTAAGCCGCACTCTATTCACACTCTTCGACGGGATTTTTTGC	4860
Db	1021	ACTAACCCTGATTTCTTATTCACATTTTTCGATGTGATTTCTGC	1062
RESULT 15			
AAC68978			
ID AAC68978 standard; DNA; 786 BP.			
XX AAC68978;			
XX 27-FEB-2001 (first entry)			
XX Merozoite surface protein-133 coding sequence.			
XX Merozoite surface protein; proteasozacide; vaccine; malaria; ss.			
XX Plasmodium falciparum.			
NN W0200063245-A2.			

```

XX 26-OCT-2000.
PD
XX 20-APR-2000; 2000MO-GB01558.
PF
XX 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-031817.
PR 25-MAY-1999; 99CA-2271451.
XX (MEDICAL RES COUNCIL.
PA
XX Holder A, Birdsal B, Feeney J, Morgan W, Syed S, Uthaiplbull C;
PI WPI: 2001-015762/02.
DR P-PSDB: AAB37610.
XX
XX Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria -
PT
XX Example 5; Fig 15; 126bp; English.
PS
XX The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1-4-2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1-4-2, compared to natural MSP-1-1-9. The non-natural variants of the
CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is the MSP-133
CC coding sequence.
XX
SQ Sequence 786 BP; 245 A; 164 C; 159 G; 218 T; 0 other;
Query Match 10 6%; Score 525.2; DB 22; Length 786;
Best Local Similarity 79.3%; Pred. No. 1.6e-117;
Matches 623; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
OY 3799 GCGTCACTCTCCGTCGATGATTAACATCTGTCACAAATCGAAGCAAGTGA 3858
DB 1 GCGGTACTCCACATCTGTATCGATACATCTCTCTAAGATTGAGAAAGATACGAGTGC 60
OY 3859 CTCTATCTGAACCTCTGCGAGGCTGTATAGTCTCTCTCAAGAAGACGTGGAATTAAC 3918
DB 61 TTGTACTTGAAGGCTCTGCGGCTGTCTACAGATCCCTGGAAGCAAGTCAAGAAACAC 120
OY 3919 GTGATACCTTCAATGTCGAACGTCGAAGACATCTGCAACGCGCTTTATAGAGAGAA 3978
DB 121 GTCTATGACTTTCACAGCTTAAAGTCGAAGACATTTTGAACCTCGAATTCACAGAGAA 180
OY 3979 AATTTCAAGACGCTTTGAGAGACGATGATTCCTTAAAGACCTGACCTCTTAAC 4038
DB 181 AACTTAAAGACGCTTGTGAGTGTGACTGTGATTCATACAGAGATTTGACTTTCTTAAC 240
OY 4039 TACGTTGCAAGGACCCATCAAGTCTCTCAATAAAGAGAGGAGTAATTTCTGTCT 4098
DB 241 TACGTTGTTAAGGACCCATCAAGTCTCTGAACAAGAGAGAGACAAGTCTGTCTGC 300
OY 4099 AGTTAACTATATCAAGAGCTCCATCGACAGCATATCAATTCCTATGATGTGCTG 4158
DB 301 TCTTAACTACCTAATTAAGGACCTCAATGATATCAATTCGCTAAGGAGCTCTG 360
OY 4159 GGGTATTACAGATCTGAGAGAAATACAGTCTGACCTTGAATTTAAAGATAT 4218
DB 361 GGTACTACAGATCTTGTCTGAGAGATACAGTCTGACTGTGATTCATCAAGAGATAC 420
OY 4219 ATCAACGATTAAGCAGAGCGAATGAATAATATCTGCCCTCTCGATTAACATCAAGAAC 4278
DB 421 ATCAATGATTAAGCAGAGAGATGAATAAGTACCTGCAATTCGATTAACATTAAGTAACT 480
OY 4279 CTGTAAAGACAGTGAAGCAGCAAAATCGACCTTGTGTAATTAACCTGGAGGCGCAAGTCT 4338
DB 4279 CTGTAAAGACAGTGAAGCAGCAAAATCGACCTTGTGTAATTAACCTGGAGGCGCAAGTCT

```

```

DB 481 TTGTACAGACGCTGTAAAGCATTAAGATCGATCTGTCTCATTTATTTGGACCAAGGTC 540
OY 4339 CTCACATTAATTAAGAGAGAGCAATGTGAGATTAAATCAAGAGCTGAACCTCTC 4398
DB 541 TTGCATTAATTAAGAGAGAGCAATGTGAGATTAAATCAAGAGCTGAACCTCTC 600
OY 4399 AAAACATTCAGAGAGAGCTGCGAGATTTTCAAGAAAAATTAATTAATTTGCGAATTTGCA 4458
DB 601 AAGACCATTCAGAGATTAAGCTGCGATTTTCAAGAGAAACAACTGCTGTGATTCGCT 660
OY 4459 GACCTGCTACGATTAATTAAGCAACAAATCTGCTGACCAAGTTTCTGTCACCTGCAATG 4518
DB 661 GATTTGTCCAGATTAATTAAGCAACAAATCTGCTGACCAAGTTTCTGTCACCTGCAATG 720
OY 4519 GTGTTGGAAGACCTGCGCAAAACAGTGTGAGCAATCTGCTGCGAGGCAAGCTGCAAGGCG 4578
DB 721 GTTTTGGAGACCTGCGTAAGTGTGCTGCTGCAACCTGTTGATGTGATTAATTTGCAAGGCT 780
OY 4579 ATGCTG 4584
DB 781 ATGCTG 786
RESULT 16
AAF89840
ID AAF89840 standard; DNA; 1128 BP.
XX
XX AAF89840;
AC
XX
XX 23-JUL-2001 (first entry)
DE
XX DNA encoding major merozoite surface protein-1 fragment of 42kDa.
XX
XX Major merozoite surface protein-1; MSP1-42; melltin signal peptide;
XX malaria vaccine; ss.
XX
XX Plasmodium falciparum.
OS
XX
XX Key Location/Qualifiers
XX CDS 1..1128 /tag=1a
XX FT /product="major merozoite surface protein-1 fragment
XX of 42kDa"
XX
XX WO200134188-A1.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000MO-US31064.
XX
XX 12-NOV-1999; 99US-0165178.
XX 01-DEC-1999; 99US-0168327.
XX 22-AUG-2000; 2000US-0226861.
XX
XX (UYHA-) UNIV HAWAII.
XX (UYCH-) UNIV CHINESE HONG KONG.
XX (QUEB-) QUEBEN EMMA FOUND.
XX
XX Hui GSN, Lap-Yin P, Ho WKK.
XX
XX WPI: 2001-335879/35.
XX P-PSDB: AAB3926.
XX
XX Producing malaria vaccine, useful for treatment or prevention of all
XX forms of malaria in humans, by expressing immunogenic merozoite protein
XX fragment in a baculovirus system -
XX
XX Example 3; Page 88-89; 95pp; English.
XX
XX The present sequence encodes a major merozoite surface protein-1
XX C-terminal fragment of 42kDa (MSP1-42). This fragment is linked to a
XX melltin signal peptide, and then expressed in a in a
XX silkworm/baculovirus system. The protein is used to prepare a

```

CC malaria vaccine, which is used to treat or prevent malaria, caused by
 CC any of the four species of Plasmodium that infect humans.
 XX Sequence 1128 BP; 487 A; 149 C; 163 G; 329 T; 0 other:

Query Match 6.5%; Score 322.8; DB 22; Length 1128;
 Best Local Similarity 58.3%; Pred. No. 3.6e-68;
 Matches 644; Conservative 0; Mismatches 397; Indels 63; Gaps 2;

```

OY 3817 ATTGATACATTTCTGCCAAAATCGAGAACGATACGAGGCTGTATCTGAAACCTCTG 3876
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19 ATGGAATATATCTCTCGAGATTTGAAAATGATGATATATATTTAAACCTTTA 78

OY 3877 GCAGCGCTATAGCTCTCAAGAAACAGCGAGATATACGATGACCTTCAATGTC 3936
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 79 GCTGAGATATATAGAGCTTTAAAAACAAATTGAAAAACATTTTACATTTATATTA 138

OY 3937 AACGTGAAGACATTCGACAGCGCTTTAATAAGAGAAAATTTGAAAGACCTCTTG 3996
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 139 AATTGAAGATATCTTAAATTCACGCTCTTAAAGAAACAAATATTTCTTACATGATTA 198

OY 3997 GAGAGCGATGATTCCTATTAAGACCTGACCTCTTACTACCTAGCTGTCAAGACCA 4056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 199 GAATCTGATTTAAGCAATTTAAACATATATCCCAATGATATCATTTATTAAGATCA 258

OY 4057 TACAAGTTCCTCAATTAAGAGAGAGGATTAATTTCTGTCTAGTACACTATATCAAG 4116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 259 TTTAATATTTGATTCGAGAACAAAAACACACTTTTAAAAAGTTTACAAATATATTA 318

OY 4117 GACTCCATCGACACCGCATATCAATTTGCTATATGATGCTGGGATTTTACAGATCTG 4176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 319 GAATCAGTAAAAATGATATTAATTTGACAGAAAGATATAGTTATATTAAGAAAGTT 378

OY 4177 AGCGAAAAATACAGCTGACCTGACTCTATTTAAAAAGTATATACGATTAAG----- 4230
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 379 TTAGGAATATTAAGATGATTTAGAAATCAATTTAAAAAGTATCAAGAAAGAAAGAG 438

OY 4231 -----CAAGC 4236
    |||||

DB 439 AAGTCCATCATCACCAACAACACCTCCGTCACCAAGAAAAACAGACGACAAAG 498

OY 4237 GAGAAATGAAAAATATCTGCCCTTCTGTAATAACATCGAAACCTGTACAGACGTGAG 4296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 499 AAGGAATGAAAGTATCTCCATTTTAAACAAACATTTGAGACCTTATCAATTAAGT 558

OY 4297 GACAAATGCAACCTCTGTATATGACGAGGCCAAGTCCCTCACTATCTTACGAG 4356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 559 AATTAATGACGATTAATTAATTAAGGCAAGATTAACGATTTGATATGTA 618

OY 4357 AAGAGCAATGTGAGATTAATCAAGAGCTGAATCACTCAAAACAATCCAGACAAG 4416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 619 AAGATGAGAGCAATGTTAAATTAACATACTAGTATTAAGACATTCATGATACAA 678

OY 4417 CTGGCAGATTTCAAGAAAAATTAACATTTCTGGAATGACAGCTGTCTACGATTA 4476
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 679 ATAGATCTTTTAAAAAACCTTACGACTTCGAAGCAATTTAAAAATGATAATGATAT 738

OY 4477 AACACACAAATTCCTGACCAAGTTTCTGTCACACTGCGATGCTTGTGAANAACCTGCC 4536
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 739 ACGAAAAAAGATATCTTGGAATAATTAATTAAGTATGATGAGATTAAG---TTCAAAATTTTCT 795

OY 4537 AAAACAGTGTGAGCAATCTGTCAGCGCAACCTGACGGGCTCAACATCTCCAG 4596
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 796 AATACAAATTAATTCAAATTAATTAAGGAAATTCGAAGATATTTAAACATTTTCACAA 855

OY 4537 CACCAATCGTGAGAGAAAGAGTCCCCAGAAATAGCGGCTGTTCAGCATCTGACGAG 4656
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 856 CACCAATCGTGAGAGAAAGAGTCCCCAGAAATTCGAAGATATTTAAACATTTTCAGAA 915

OY 4657 CGCGAAGGTGCAAGTGTCTCTGACATCAACAAGAGAGATTAAGTGGCTGAGAGAC 4716
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 916 AGAGAGAGATTAATGATTTTAAATTAACAACAAGAGGATGATTAATGATGTAAT 975
  
```

```

OY 4717 CCAACCTACCTGCATGAATGAANAACATGCGGGTGTGAGCGCGCATTAATGACCGAG 4776
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 976 CCAATTCCTACTTGTAGAGAAATATATGATGATGATGATGATGATGATGATGATGAT 1035

OY 4777 GAGACAGCGGCTCTAACGAGAAAGAAATACATCTGAGTGTACTTAAGCCGACTCTAT 4836
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1036 GAGATTCAGGTAGAGCAGAGAAAGAAATACATCTGATGATGATGATGATGATGATGAT 1095

OY 4837 CCACCTTCGACGGCATTTTTCG 4860
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1096 CCACCTTCGATGATGATTTTTCG 1119
  
```

RESULT 17

AA81148
 ID AA81148 standard; DNA; 660 BP.

XX AA81148;

XX 11-NOV-1990 (first entry)

XX DNA sequence encoding polypeptide p190-1.

XX Polypeptide p190-1; P.falciparum; merozoite; vaccine.

XX Plasmodium falciparum.

XX OS

XX FH

XX FT

XX CDS

XX Key

XX Location/Qualifiers

XX 1..660

XX /*tag= a

XX EP283829-A.

XX 28-SEP-1988.

XX 08-MAR-1988;

XX 88EP-0103564.

XX 19-MAR-1987;

XX 87GB-0006599.

XX (HOFF) F. HOFFMANN-LA ROCHE & CO.

XX Ulrich C, Gentz H, Takacs B;

XX WPI; 1988-272339/39.

XX P-PSDB; AAP80544.

XX New immunogenic polypeptides - derived from plasmodium falciparum 190 kD

XX surface antigen precursor, useful in vaccines, and encoding DNA sequences

XX Disclosure; ; pp; German.

XX The DNA encodes a region of a polypeptide. The region contains at least

XX one epitope of the 190KD precursor of the major merozoite surface antigen

XX of P.falciparum.

XX Sequence 660 BP; 281 A; 84 C; 102 G; 193 T; 0 other;

XX

Query Match 6.3%; Score 308.8; DB 9; Length 660;

Best Local Similarity 74.1%; Pred. No. 7.1e-65;

Matches 391; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```

OY 475 ACACGTGTGATTAACATGCTGCTCAAAATCTGATTTGACGTTAGAGAGATCAAT 534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 25 ACTTGTGTGATTAATTCATCATGCTTCAAAATTTAATTTGATGATGATGATGATGATGAT 84

OY 535 GAATCTGTATACAGTTGAATTTCTAGCTTCGACTTCTAAGGCCAAACTGAATGAGTT 594
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 85 GAATATATATATTAATTAACCTTTATTTGATTAATTAAGGCAAAATTAATTAATGTA 144

OY 555 TGCGCAATGACTATTTGCAAAATTCATTTGAAGATCAAGCCACAGAGTTGAC 654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 145 TGTGCTAATGATTAATTTGCAAAATACCTTTCAATTTAAATTCGCAAAATGAATTAAGAC 204
  
```

QY	655	GTTTGAAGAGTTGGCTTGGATATTCGCAAGCCTCGCAACATCAAGGACAAATGTG	714
Db	205	GTACTTAAAAAACTTGtGTTCGGATATGAAAAACATTAGCATATTAAATAAATGTA	264
QY	715	GGAAAGATGGAAGATTATATATTAATAAAGAAATAGCAAGACCATCGAATTAACGAGCTG	774
Db	265	GGAAAAAATGGAAGATTACATTATAAAAAAATATAAAAAACCATAGAAATATTAATGAATTGA	324
QY	775	ATCGAAGATCCAAAAAGACCATTAGACAAAAATTAAGATGCAACCAAGAGAGAAAGAAAG	834
Db	325	ATTGAGAGAAATAGAAAAACAATTGATATAAAATTAAGATGCAACTTAAGAAAGAAAGAAAA	384
QY	835	AAGAAATTGACCAAGGCCAGTACGACCTGTCCATCTATATAAACAACGTTGAAGAAAGCC	894
Db	385	AAAAAATATATCCAAAGCTCATATGATCTCTTTTATTATACATAAACAATTAGAAAGAGCA	444
QY	895	CATAACCGATCAGCGTACTGTGGAAGAGCGCTATACACACCTCAAGAAGAAATGAAATATC	954
Db	445	CATATATTAAATTAACCGTTTATGAAAAAACGATTACACACTTTAAAAAATAATGAAAAACATT	504
QY	955	AAAGAACTGTCGACCAAGATTAATGAATTAAGAAATCTCGCGCCAGCC	1002
Db	505	AAGGATTACTTGATTAAGATTAATGAATTAATTAATATCCCGACCGGCGC	552

RESULT 18
AAN81151
ID AAN81151 standard; DNA; 1068 BP.

AC AAN81151;

DT 11-NOV-1990 (first entry)

DNA sequence encoding polypeptide p190-3.

KW Polypeptide p190-3; *P. falciparum*; merozoite; vaccine

OS *Plasmodium falciparum*.

FH	Key	Location/Qualifiers
1000		

FT / *tag= a

PN EP283829-A.

PD 28-SEP-1988

PF 08-MAR-1988; 88EP-0103564.

PR 19-MAR-1987; 87GB-0006599.

PA (HOFF) F. HOFFMANN-LA ROCHE & CO.

Ulrich C, Gentz H, Takacs B;
PI
xy

DR WPI; 1988-272339/39.

XX
XX
Nov 4 1966

PT surface antigen precursor, useful in vaccines, and encoding DNA sequences

PS Disclosure; ; 5pp; German.
XY

CC The DNA encodes a region of a polypeptide. The region contains at least

CC of P. falciparum.

Sequence 1068 BP; 471 A; 129 C; 161 G; 307 T; 0 other;

Query Match	6.3%	Score 308.8	DB 9	Length 1068
Best Local Similarity	74.1%	Pred. No. 9e-65		
Matches 391	Conservative	0	Mismatches 137	Indels 0
				Gaps 0

OY	475	ACACGTCGTATACATTCATTCAGCGCTTCAAAATATCGATTCGATTCACGAAGATCAAT	534
Db	433	ACTTTGTCTGATTAATTCATTCAGCTTCGTTCAAAATTTAATTCATTCGATTCGATTCGAAGAAATTAAT	432
OY	535	GAACCTCTGTACAGTTGAAATTTCTTACTTGCGACTTGTCTAAGGGCCAAACTGAATGACGTT	594
Db	493	GAATTAATTAATTAATTAATTAACCTTTATTTTGGATTTATTAAGAGCAAAATTAATTAATGTA	552
OY	595	TGGCGCATGACATTCGTTCGCAATTCACATTCGATTTGAAGATCACAGCAACGAGTTGAC	654
Db	553	TGTGCTATATGATTAATTCGCAATTCACCTTTTCATCTTTAAATTCGTGCAAAATGAAATTTAGAC	612
OY	655	GTATTTGAAGAAGTTGGCTTCCTGGATATTCGCAAGCCCTCTCGACACATCAAGGACATATGTG	714
Db	613	GTACTTAAAAAACTGTGGTTCGATATAGAAAAACCATTAAGACATTTAAGAATTAATGTGA	672
OY	715	GGAAAGATGGAAAGATTTATTTAAAAAGAAATTAAGAAAGCAATCCGAGCAATTAACGACCTG	774
Db	673	GGAAAAATGGAGAGTTATTCATTAAAAAAAATTAAAAAAACCATAGAAATTTAATTAAGAAATTA	732
OY	775	ATCGACAGATATCAAAAAGACCATAGACAAAATTAAGATGCAACCAAGAGAGAAAG	834
Db	733	ATTGAGAAAGATTAAGAAACAATTTGATTAATAATTAAGATGCAACTTAAGAAGAAAGAAAA	792
OY	835	AAGAAGTTGTACAGGCCCACTACGACCTGTGTCATCTATACAAACAGCTTGAAGAAAGCC	894
Db	793	AAAAAATTAATCCAAAGCTCAATATGATCTTTTATTTTACATAATAACATTTAGAGAAAGCA	852
OY	895	CATTAACCTCATACACGCTACTGAGACAGCGCATATACACCCCTCAAGAAAGATGAAAATATC	954
Db	853	CATATTTAATTAAGCGTTTGTGAAAAACGATTTGACACTTTAAAAAAAATGAAAACATT	912
OY	955	AAAGACATGCTCGACAGATTAATGAATTTAAGATTCCTCCGCGACGC	1002
Db	913	AAGGATTAATCTGATAGATTAAGAAATTAATAATCCCCACCGCGCG	960

RESULT 19	
AAT80403	
ID AAT80403 standard; DNA;	354 BP.

AC AAT80403;

DT 25-MAR-1998 (first entry)

DE PFMSP1(p19)A coding sequence.

KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;

XX

OS Synthetic.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

$$FT \quad \frac{/*tag=a}{a}$$

FT /note="sequence contains 2 in frame stop codons at

FT	misc_feature	1..285	
----	--------------	--------	--

	/note= "derived from P. falciparum MSP1 p19 sequence"
--	---

$$F_T = \frac{1}{c} \frac{d}{dt} \int_{\text{tag}} \dots$$

XX

XX

XX

XX
XX

XX (INSP) INST PASTEUR.
 PA (UYNV) UNIV NEW YORK STATE.
 XX Barnwell JM, Longacre-Andre S, Mendis K, Nato F;
 PI Roth C;
 XX WPI: 1997-425034/39.
 DR P-PSDB: AAM22592.
 XX
 PT Recombinant protein containing Plasmodium mezoizote surface
 PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
 PT antibodies for diagnosis and protein purification
 XX
 PS Disclosure; Fig 1B: 79pp; French.
 XX
 CC This is the nucleotide sequence encoding a recombinant protein comprising
 CC nucleotides encoding amino acids 1613-1705 of the Plasmodium falciparum
 CC mezoizote surface protein 1 (MSPI) 19 kD C-terminal fragment (p19),
 CC linked to a sequence encoding a glycosylphosphatidylinositol membrane
 CC anchoring sequence. p19 is the C-terminal fragment of the 42 kD MSPI
 CC from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX
 SQ Sequence 354 BP; 99 A; 92 C; 88 G; 75 T; 0 other;

Query Match 5.0%; Score 246.6; DB 18; Length 354;
 Best Local Similarity 81.7%; Pred. No. 7.1e-50;
 Matches 285; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 4583 TGACATCTCCAGCAGCAATGCGTGAAGAAAGTGGCCCCAGAAATAGCGGCTTTCA 4642
 DB 5 TCAACATCTCCAGCAGCAATGCGTGAAGAAAGTGGCCCCAGAAATAGCGGCTTTCA 64
 QY 4643 GGCACTTGAGAGAGCGGAGAGAGTGTCTCTGCACTACAAACAGAGAGATA 4702
 DB 65 GACACTTGAGAGAGAGAGAGAGTGAATGTCTGCACTACAAACAGAGAGATA 124
 QY 4703 AGTGGTGGAGAAACCCCTACCTGCAATGAAACAAATGGGGGTGTACGCCGATG 4762
 DB 125 AGTGGTGGAGAAACCCCTACCTGCAATGAAACAAATGGGGGTGTACGCCGATG 184
 QY 4763 CTAATGACACGAGAGAGAGAGCGGCTCTAAGCAAGAAATGACATGAGTACTA 4822
 DB 185 CCAATGACACGAGAGAGAGAGCGGCTCTAAGCAAGAAATGACATGAGTACTA 244
 QY 4823 AGCCGACTCTATCCACTCTTTCGACGAGGATTTTGTCCAGCTCTAATTTCTGGGCA 4882
 DB 245 AACCGACTCTATCCACTCTTTCGACGAGGATTTTGTCCAGCTCTAATTTCTGGGCA 304
 QY 4883 TCTCTTCTGCTGATCTCTATGCTGATCTGTACAGCTTCACTAATA 4931
 DB 305 TCTCTTCTGCTGATCTCTATGCTGATCTGTACAGCTTCACTAATA 353
 RESULT 20
 AAT94550
 ID AAT94550 standard; DNA: 354 BP.
 XX AAT94550:
 XX 25-MAR-1998 (first entry)
 XX
 DE PfMSPI(p19)A coding sequence.
 XX
 KW Chimeric: Plasmodium vivax; mezoizote surface protein: MSPI. p19;
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
 OS Plasmodium falciparum.
 XX Synthetic.
 XX

FH Key Location/Qualifiers
 FT CDS 1..354
 FT /*tag= a
 FT /product= "PfMSPI(p19)A with GPI anchoring sequence"
 FT /note= "sequence contains 2 in frame stop codons at
 FT the 3'-end"
 FT misc_feature 1..285
 FT /*tag= b
 FT /note= "derived from P. falciparum MSPI p19 sequence"
 FT misc_feature 286..354
 FT /*tag= c
 FT /note= "glycosylphosphatidylinositol anchoring sequence"
 PN M09730158-A2.
 PD 21-AUG-1997.
 XX
 XX 14-FEB-1997; 97WO-FR00290.
 XX
 XX 14-FEB-1996; 96FR-0001822.
 PA (INSP) INST PASTEUR.
 PA (UYNV) UNIV NEW YORK STATE.
 PI Barnwell JM, Longacre-Andre S, Mendis K, Nato F;
 PI Roth C;
 XX
 XX WPI: 1997-425033/39.
 DR P-PSDB: AAM36103.
 XX
 PT Recombinant protein containing the mezoizote surface protein-1 p19
 PT fragment - useful in anti-malarial vaccines, diagnosis and protein
 PT purification
 XX
 PS Disclosure; Fig 1B: 85pp; French.
 XX
 CC This is the nucleotide sequence encoding a recombinant protein comprising
 CC nucleotides encoding amino acids 1613-1705 of the Plasmodium falciparum
 CC mezoizote surface protein 1 (MSPI) 19 kD C-terminal fragment (p19),
 CC linked to a sequence encoding a glycosylphosphatidylinositol membrane
 CC anchoring sequence. p19 is the C-terminal fragment of the 42 kD MSPI
 CC from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX
 SQ Sequence 354 BP; 99 A; 92 C; 88 G; 75 T; 0 other;

Query Match 5.0%; Score 246.6; DB 18; Length 354;
 Best Local Similarity 81.7%; Pred. No. 7.1e-50;
 Matches 285; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 4583 TGACATCTCCAGCAGCAATGCGTGAAGAAAGTGGCCCCAGAAATAGCGGCTTTCA 4642
 DB 5 TCAACATCTCCAGCAGCAATGCGTGAAGAAAGTGGCCCCAGAAATAGCGGCTTTCA 64
 QY 4643 GGCACTTGAGAGAGCGGAGAGAGTGTCTCTGCACTACAAACAGAGAGATA 4702
 DB 65 GACACTTGAGAGAGAGAGAGTGAATGTCTGCACTACAAACAGAGAGATA 124
 QY 4703 AGTGGTGGAGAAACCCCTACCTGCAATGAAACAAATGGGGGTGTACGCCGATG 4762
 DB 125 AGTGGTGGAGAAACCCCTACCTGCAATGAAACAAATGGGGGTGTACGCCGATG 184
 QY 4763 CTAATGACACGAGAGAGAGCGGCTCTAAGCAAGAAATGACATGAGTACTA 4822
 DB 185 CCAATGACACGAGAGAGAGCGGCTCTAAGCAAGAAATGACATGAGTACTA 244
 QY 4823 AGCCGACTCTATCCACTCTTTCGACGAGGATTTTGTCCAGCTCTAATTTCTGGGCA 4882
 DB 245 AACCGACTCTATCCACTCTTTCGACGAGGATTTTGTCCAGCTCTAATTTCTGGGCA 304
 QY 4883 TCTCTTCTGCTGATCTCTATGCTGATCTGTACAGCTTCACTAATA 4931

QY	337	AGAACCAATCCATCTGACACACCTAGGATTCGAGCCCAAGCTCTAGCGCAGCTCAAG	396
Db	133	AGTACCAATCTCTCAATATATTCAGATGATCATATCTTAATCTTAACGCTGATTTAAA	192
QY	397	CACCGAGTGAGAACTATCTCTCACTATCCAGAGCTGAAGTACCACAGTTGTTCGAC	456
Db	193	CATAGAGTTCAAAATTAATCTGTTCCACTATTAAAGAACTCAAAATATCCGAACCTTTGAT	252
QY	457	CTCAGTATCATATGCTGACACTGTGTGATTAACATTCATGGCTTCAAAATATCTGATTGAC	516
Db	253	TTAACCAATCATATGTTAACTTTGTGTGATTAATATCATGTGTTCAAAATATTAATTGAT	312
QY	517	GGTTACGAAGAGATTCATAGAACCTCGTACAGTTCGAATTTCTACTTCGACTTGCTAAG	576
Db	313	GGATATGAAGAAATTAATGATTAATTAATTAATTAACCTTTTAATTAATGATTATTAAGA	372
QY	577	GCCAAACTGAATGACGTTTGGCCCAATGACTATTTGTCMAATTCATTCAAATTTGAAGATC	636
Db	373	GCCAAATTAATATGATGATGATGTCGCAATAGTAAATTTGTCMAATTAACCTTTAAATTT	432
QY	637	AGAGCCACAGAGTGTGACCGTATTGAAGAAGTTGGTCTTCGGATATTCGAACGCTCTGCAC	696
Db	433	CTGCGCAAAATGAATTAACGCTACTTAAAAAAATTTGTGTCGATATTAAGAAAAACCATTTAGAC	492
QY	697	AACATCAAGGACATGCTGGGAAGATGGAAGATTATTTAAAAAGAAATTAAGAAGCAGATC	756
Db	493	AATATTAAAGATATATGTAGCAAAATGGAAGATTACTTTAAAAAAATTAACAACCATTA	552
QY	757	GAGAACATTATACGAGCTGATCGAAGAATCCMAAAGACCATAGACAATAAATTAAGATGCA	816
Db	553	GCAAATATTAATGAATTAATTAATGAAGAAAGTAAAGAAACAAATTAATGATTCGGGCAACGTTGTT	612
QY	817	ACCAAGAGAGGAG 829	
Db	613	GCCATTGCTGCAG 625	
RESULT 23			
AAN81150			
ID	AAN81150 standard; DNA; 456 BP.		
XX	AAN81150;		
AC			
XX			
DT	11-NOV-1990 (first entry)		
DE	DNA sequence encoding polypeptide p190-2b.		
XX			
KW	polypeptide p190-2b; P.falciparum; merozoite; vaccine.		
XX			
OS	Plasmodium falciparum.		
XX			
Key	Location/Qualifiers		
FT	1..456		
FT	/tag= a		
XX			
PN	EP283829-A.		
XX			
PD	28-SEP-1988.		
XX			
PF	08-MAR-1988; 88BP-0103564.		
XX			
PR	19-MAR-1987; 87GB-0006599.		
XX			
PA	(HOFF) F HOFFMANN-LA ROCHE & CO.		
XX			
PI	Ulrich C, Gentz H, Takacs B;		
XX			
DR	WPI; 1988-272339/39.		
XX			
DR	P-PSDB; AAP80546.		
XX			
PT	New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD surface antigenic precursor, useful in vaccines, and encoding DNA sequences		

Query Match	4.1%	Score 201.8	DB 9	Length 456
Best Local Similarity	67.5%	Pred. No. 6.5e-39		
Matches 284	Conservative 0	Mismatches 137	Indels 0	Gaps 0
Sequence 456 BP	200 A	63 C	70 G	123 T
0 other:				
Query	3214	GCACGATGCGCGAGACAGAGACACTCTGGAGAACACCAAGATTCTTCTCAACACTAC	3273	
Db	16	GCTGAATATGCAAACTGAAAACACATTAGAAAACAAATATTATTGAAACATTAT	75	
QY	3274	AAAGCCTCGTCAATATTATAATGCGAGCTCTTCTCTGTAAGACTCTCCGAGAG	3333	
Db	76	AAAGCACTTGTTAAATATTATAATGTTGTAATCATCTCCATTAAACCTTTAAGTGAAGAA	135	
QY	3334	AGCATCCAGACGAGGATTAATCTACGACGCTCGAGAACTTCAAGCTCGTCAAGCTC	3393	
Db	136	TCAATTTCAACACAGACATATTATGTCAGTTTAGAATACTTTAAGATTAAAGTAAATTA	195	
QY	3394	GAAGCAAGCTGAGAGACCACTGAACCTGAGAGAGAAAGCTCAGCTCCTCTAGC	3453	
Db	196	GAAGCAAAATTTAAGATATTATTAAATTTTAGAAGAAATTTATCATCTATTATCAAGA	255	
QY	3454	GGACGTCATCACCTGATGCGCCGACCTCAAGAACTATTAGAACAACTACACCGGC	3513	
Db	256	GGTTTACATCATTTAATTTGACAGAAATTTAAAGAAAGTATATAAAATTTATACAGGT	315	
QY	3514	AATAGCCCAAGGAGATATATACAGACGTGAATACGACACTGGAATCTTACAGAAAGTTC	3573	
Db	316	AATTTCTCCAAAGGCTAATATATATACGATGTTAACAAATGCAATTACATTTCAAAAAATTT	375	
QY	3574	GTGCTGAAGAGACAGATGTGCGCCACTGTGTGTCTGTAATCTGTGCTCGACACACTGGAG	3633	
Db	376	CTCCCGAAGAAACAGATGTGTGCAACAGTTGTAAAGTGAAGATGATCCGTGACCTGACG	435	
QY	3634	C 3634		
Db	436	C 436		
RESULT 24				
AA81149				
ID	AA81149	standard: DNA; 462 BP.		
AC	AA81149			
DT	11-NOV-1990	(first entry)		
DE	DNA sequence encoding polypeptide p190-2a.			
KW	Polypeptide p190-2a; P.falciparum; merozoite; vaccine.			
OS	Plasmodium falciparum.			
XX	Key	Location/Qualifiers		
FT	CDS	1..462		
XX		/tag= a		
XX	EP283829-A.			
XX	28-SEP-1988.			
XX	08-MAR-1988;	88EP-0103564.		
XX	19-MAR-1987;	87GB-0006599.		
XX	(HOFF) F. HOFMANN-LA ROCHE & CO.			


```

XX  Ulrich C, Gentz H, Takacs B;
PI  WPI: 1988-272339/39.
XX  P-PSDB; AAB80545.
DR  New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
XX  surface antigen precursor, useful in vaccines, and encoding DNA sequences
XX  PS  Disclosure: ; pp; German.
XX  The DNA encodes a region of a polypeptide. The region contains at least
CC  one epitope of the 190kD precursor of the major merozoite surface antigen
CC  of P.falciparum.
XX  Sequence 462 BP; 200 A; 65 C; 72 G; 125 T; 0 other;

Query Match      4.1%; Score 201.8; DB 9; Length 462;
Best Local Similarity 67.5%; Pred. No. 6.6e-39;
Matches 284; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY  3214 GCCGAGATCCGCGAGACAGAGAACACTCTGAGAACACCAAGATTCTTCTCAACACTAC 3273
DB  22 GCTGAATACGAGAACTGAAAACACATGTAGAAAACAAAATATATATTGAAACATTAT 81
QY  3274 AAAGCCCTGTCAGATATATATATGCGAGCTCTCTCTGAAAGACTCTCTCCGAGAG 3333
DB  82 AAAGCACTTGTATATATATATATGATGATCATCTCCATTAATAAACTTTAAGTAGAAG 141
QY  3334 AGCATCCAGACCGAGATACCTACGCGCTGAGAACCTCAAGTCCGTCTAACCTC 3393
DB  142 TCATTTCAACAGAGATATATATGCGAGTTTAGAAAACCTTTAAAGTATTAAGTAA 201
QY  3394 GAAGGCAAGCTGAGAGCAACACTGAACTGAGAGAGAGAGAGCTCAGCTACTCTAGC 3453
DB  202 GAAGGAAATTAAGGATATATATATTAATTAAGAAAAGAAAATATATATCTTATCAGA 261
QY  3454 GGAAGCTGATCACTGATGCGGAGCTGAGAGAGTCAATTAAAGCAAGAACTACACGGC 3513
DB  262 GGTTCATCATCTTAAATGCTGATTAAGAAAGTAAATAAAATTAATTAACAGGT 321
QY  3514 AAATAGCCCAAGCGAGATATATACAGAGTGAATTAAGCAGTGAATTTCAAGAGTTC 3573
DB  322 AATTTCTCAAGCGTAAATATATACGATGTTAATCATGATTAATCTTCAAAAATTT 381
QY  3574 CTGCTGAAAGAGACAGATGTCGCCACTGTGCTGCTGATCTGCTCCGACACACTGAG 3633
DB  382 CTCACGAAAGAGACAGATGTTGCAACAGTGTAAAGTGAAGTGAATGCGTGCAGCTG 441
QY  3634 C 3634
DB  442 C 442

RESULT 25
AAC68977
ID  AAC68977 standard; DNA: 333 BP.
XX  AAC68977;
XX  AAC68977;
DT  27-FEB-2001 (first entry)
XX  Merozoite surface protein-119 coding sequence.
XX  Merozoite surface protein; protazoacide; vaccine; malaria; ss.
OS  Plasmodium falciparum.
XX  WO200063245-A2.
XX  26-OCT-2000.
XX  20-APR-2000; 2000WO-GB01558.

```

```

XX  20-APR-1999; 99GB-0009072.
PR  13-MAY-1999; 99US-031817.
PR  25-MAY-1999; 99CA-2271451.
XX  (MED-) MEDICAL RES COUNCIL.
XX  Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiybull C;
XX  WPI: 2001-015762/02.
XX  P-PSDB; AAB37609.
DR  Novel variants of the C-terminal fragment of Plasmodium merozoite
PT  surface protein-1, useful as vaccines for treating or preventing
PT  malaria -
XX  Example 5; Fig 15; 126pp; English.
XX  The present invention relates to non-natural variants of a C-terminal
CC  fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC  non-natural variants have reduced affinity for at least 1 antibody
CC  capable of blocking a second antibody that inhibits the proteolytic
CC  cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC  one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC  MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
CC  present invention are useful for immunizing a mammal against malaria, and
CC  can be used to treat malaria. The present sequence is the MSP-119
XX  coding sequence.

QY  Sequence 333 BP; 107 A; 75 C; 73 G; 78 T; 0 other;
Best Local Similarity 4.0%; Score 198.4; DB 22; Length 333;
Matches 241; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY  4668 ACCGCGAGGCGATGCTAATCTCCAGCAGCAAGTCGTAAGAAACAGTCCCGCAGA 4627
DB  20 ACATTGAAGGTAGACACACACTTGCCCAACACCATGCTTAAGAACATGTCACAAA 79
QY  4628 ATAGCGGCTGTTTCAGGCACTGTGAGCAGCGCGAAGAGTGCATGTCCTGTAATACA 4687
DB  80 ACTCCGATGTTTCAGACACTCTGAGCAGAGAGAGAAAGTAAGTGTCTGTGAATACA 139
QY  4688 AACAGAGAGAGATAGTGGTGTGAGAACCCAAACCTTACTGATGAAGAAACATGGCG 4747
DB  140 AGCGAGAAAGGTGAATAGTGTGAGAACCCAAACCTTACTGATGAAGAAACATGGCG 199
QY  4748 GGTGTGACGCGGATGCTAATATGACACGAGAGAGACGCGCTTAAGCGAAAGAAATCA 4807
DB  200 GATGCGAGCGCTGACGCTAAGTGCACCGAAGAGACTGTGTTCAACGGAAGAGATT 259
QY  4808 CATGCGAGTGTACTAAGCCGCGACTCTATCCACTCTTGCACGCGGATTTTGTCCAGCT 4867
DB  260 CTTCGGAATGTACTAAGCCGAGACTCTTACCTTTGTTGATGATGATCTCTGTTCTCCT 319
QY  4868 CTAAATTTCTGCG 4879
DB  320 CTAACTAAGTGG 331

RESULT 26
AAT80404
ID  AAT80404 standard; DNA: 387 BP.
XX  AAT80404;
XX  AAT80404;
DT  25-MAR-1998 (first entry)
XX  PFMSP1(p19)s coding sequence.
XX  PFMSP1(p19)s coding sequence.
XX  Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW  Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX  Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.

```

```

OS Chimeric - Plasmodium vivax.
XX Chimeric - Plasmodium falciparum.
FH Key Location/Qualifiers
FT CDS 1..387
FT /tag= a
FT /product= PfMSP1(p19)S
FT /note= "sequence contains 2 in frame stop codons at
FT the 3' end"
FT sig_peptide 1..57
FT /tag= b
FT mat_peptide 58..381
FT /tag= c
FT misc_feature 1..96
FT /tag= d
FT /note= "sequence derived from P. vivax MSP1"
FT misc_feature 97..102
FT /tag= e
FT /note= "sequence derived from generated restriction
FT site"
FT misc_feature 103..381
FT /tag= e
FT /note= "sequence derived from P. falciparum p19 coding
FT sequence"
FT W09730159-A2.
XX 21-AUG-1997.
XX 14-FEB-1997; 97WO-FR00291.
XX 14-FEB-1996; 96FR-0001821.
XX (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX Barwell JW, Longacre-Andre S, Mendis K, Nato F;
XX Roth C;
XX WPI: 1997-425034/39.
XX DR P-PSDB: AAW22592.
XX Recombinant protein containing Plasmodium merozoite surface
XX protein-1 p42 fragment - useful in antimalarial vaccines, also new
XX antibodies for diagnosis and protein purification
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the nucleotide sequence encoding a chimeric protein comprising
XX nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
XX surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
XX 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
XX CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX CC The recombinant protein can be used for the production of anti-malarial
XX CC vaccines, where the p19 fragment provides a high level of protective
XX CC immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other:
SQ
Query Match 3.9%; Score 194.8; DB 18; Length 387;
Best Local Similarity 81.3%; Pred. No. 3e-37;
Matches 226; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```

```

DB 221 AGTGGGTGAGAAACCCCAACCGACCTGTACGAGAACAAGCGGCTGTGACGACAG 280
QY 4763 CTAAATGCACCGAGAGACAGCGGCTTAAGGAAAGAAATACATCGAGTACTA 4822
DB 281 CCAAAATGCACCGAGAGAGACTCGGAGCAGCAAGCAAAATCAGTGTGAGTACCA 340
QY 4823 AGCCGACCTCTATCCACTCTCTGACGAGATTTTTC 4860
DB 341 AACCGACTGTACCCGCTGTTCGACGACATCTTCTGC 378

RESULT 27
AAT94549
ID AAT94549 standard; DNA; 387 BP.
XX
XX AAT94549;
XX
XX 25-MAR-1998 (first entry)
XX
XX PfMSP1(p19)S coding sequence.
XX
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX
XX Chimeric - Plasmodium vivax.
XX OS Chimeric - Plasmodium falciparum.
XX
XX Key Location/Qualifiers
XX FH CDS 1..387
XX FT /tag= a
XX FT /product= PfMSP1(p19)S
XX FT /note= "sequence contains 2 in frame stop codons at
XX the 3' end"
XX sig_peptide 1..57
XX FT /tag= b
XX FT mat_peptide 58..381
XX FT /tag= c
XX FT misc_feature 1..96
XX FT /tag= d
XX FT /note= "sequence derived from P. vivax MSP1"
XX FT misc_feature 97..102
XX FT /tag= e
XX FT /note= "sequence derived from generated restriction
XX site"
XX FT misc_feature 103..381
XX FT /tag= e
XX FT /note= "sequence derived from P. falciparum p19 coding
XX sequence"
XX
XX W09730158-A2.
XX 21-AUG-1997.
XX 14-FEB-1997; 97WO-FR00290.
XX 14-FEB-1996; 96FR-0001822.
XX (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX Barwell JW, Longacre-Andre S, Mendis K, Nato F;
XX Roth C;
XX WPI: 1997-425033/39.
XX DR P-PSDB: AAW36102.
XX
XX Recombinant protein containing the merozoite surface protein-1 p19
XX fragment - useful in anti-malarial vaccines, diagnosis and protein
XX purification
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the nucleotide sequence encoding a chimeric protein comprising

```

CC nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
CC surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
CC 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
X3
X4 Sequence 387 BP: 116 A; 94 C; 96 G; 81 T; 0 other;

Query Match	3.9%	Score	194.8	DB	18	Length	387
Best Local Similarity	81.3%	Pred. NO.	3e-37				
Matches	226	Conservative	0	Mismatches	52	Indels	0
						Gaps	0

QY	4583	TGAACATCTCCAGACACCATGGCTAAAGAAACAGTGGCCCGAATATCGGCTGTTC	4642
Db	101	TCAACATCTCGACACACCATGGCTAAAAAACATGTCCGGAACCTGGCTGTTC	160
QY	4663	GGCATCTGGACGAGCGGAGAGCTGCAGTGTCTCTGAATCAACAAAGAGGAT	4702
Db	161	GACACTTGGAGAGAGAGAGAGTGTAAATGTCTGTGAATCAACACAGGGGGCGACA	220
QY	4703	AGTACGTGGAGAACCCCAACCCCTTACCTGCACATGAACAAACATGGCGGGTGGACGGCGAT	4762
Db	221	AGTACGTGGAGAACCCCAACCCCACTGTATACGAGAACACGGCGCTTGGACGAGACG	280
QY	4763	CTAAATGCACCGAGAGACAGCGGCTTATACGGAAGAAAAATCAATCGAGTGTACT	4822
Db	281	CCAANTGCACCGAGAGAGAGTCCGGCGACGAACGGCAAGAAAAATCAAGTGTATACCA	340
QY	4823	AGCCGACTCTATCCCACTTTTGACAGGGATTTTTTGC	4860
Db	341	AACCGACTGTATCCGCTTTTGACGGCATTTTTCGC	378

RESULT	28
AAD22459	
ID	AAD22459 standard; DNA; 309 BP.

AC	AAD22459;
XX	
DT	12-FEB-2002 (first entry)

DE Plasmodium falciparum PfMSPI.19 insert in pUC105-01 vector

KM pUC105-01 vector; immunostimulant; vaccine; immunisation; therapeutic;
KM C3d; immune response; ds.

OS Plasmodium falciparum.

PN WO2001.77324-A1.

PD 18-OCT-2001

PF 09-APR-2001; 2001WO-GB01599.

PR 08-APR-2000; 2000GB-0008582.

PA (ADPR-) ADPROTECH LTD.

PI Steward M, Cox VF;

DR WPI; 2002-010909/01.

PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein -
XX
PS Example 13; Page 65; 87pp; English.

CC The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which
CC by virtue of third base redundancy and other variations permissible

CC within amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to elicit one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal
CC by administering the pharmaceutical composition into the human or animal
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmidum falciparum PfMSp1.19 insert
CC In pUC105-01 vector

Query Match	3.9%	Score 194.6	DB 24	Length 309
Best Local Similarity	79.6%	Pred. No. 3e-37		
Matches 230	0	Mismatches 59	Indels 0	Gaps 0

Qy	4585	ANCTATCCGAGCACCAATGTCGTGAAGAAACATGCCCCCAATATAGCGGCTGTTTCAGG	4644
Db	7	AACATTCGCCAACACCAATGCGTTTAAGAAAGCATGTCCACAAAACCTCGGATGTTTCAGA	66
Qy	4645	CATCTGGACGAGCGCCGGAAGTGCCAAAGTGTCTCTGACTACAAAACAGAGAGATTAAG	4704
Db	67	CATCTGGACGAGAGAGAAAGATGATGAATGTCTGTTCGTAACACACAGGAAGAGTATTAAG	126
Qy	4705	TGGCTGGAGAACCCAAACCTTACTCCCAATGAAACAAATGAGGGGGGTGAGCCGATCT	4764
Db	127	TGTATTAGAACCCAAACCTTACTCTGTAAACGAAACAAACGGTGGATGCGACCTTACCT	186
Qy	4765	AAATGACCGAGAGACACACGGCTCTAAACGAAAGAAATACATGCGAGTGTACTAAG	4824
Db	187	AAATGACCGAGAGAGACTGTGGTCTTAAACGAAAGAAATACATGCGAGTGTACTAAG	246
Qy	4825	CCGACATCTATCCACTCTTCACACGGGATTTTTTGTCTCCAGCTTAAT	4873
Db	247	CCGACATCTTACCTTTGTTGATGGAATCTTCTGTCTCTCTTAAT	295

RESULT 29	
AAD22460	
ID AAD22460 standard; DNA; 3147 BP.	

AC AAD22460;

DT 12-FEB-2002 (first entry)

DE PFMSPI.19-human C3d3 DNA coding sequence from pVK104-01 vector.

KW Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;

XX

OS Chimeric - Plasmodium falciparum

XX

FT	CDS	1..3147
----	-----	---------

FT /product= "Human C3d3-PfMSP1.19 protein"

$$\frac{\partial}{\partial t} F_T$$

```
FT
/*tag= | b
```

FT / *tag = | c

FT	misc_feature	73..369
----	--------------	---------

FT /note= "Antigen sequence"

PN WO200177324-A1.

XX

PD 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-GB01599.
XX
XX 08-APR-2000; 2000GB-0008582.
XX
XX (ADPR-) ADPROTECH LTD.
XX
XX Steward M, Cox VF;
XX
XX WPI: 2002-010909/01.
XX
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein -
XX
XX Claim 16; Page 66-68; 87pp; English.
XX
XX The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSPI.19-human
CC C3d3 DNA coding sequence from vector pVK104-01.
XX
SQ Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other;

Query Match 3.9%; Score 194.6; DB 24; Length 3147;
Best Local Similarity 79.6%; Pred. No. 9.9e-37;
Matches 230; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 4585 AACATCTCCAGCACCATGCTGTAAGAAAGACGCCCCCAATATAGCGCTTTTCACG 4644
DB 73 AACATGCCCCAACCAACCGCTTAAGAAAGACGCAATGTCCTCAACCTCCGATGTTCA 132
OY 4645 CATCTGGAGACGCGGAGAGTGTCTCTGTAACATACCAAGAGAGATTAAG 4704
DB 133 CATCTGGAGAGAGAGAGATGTCTCTGTAACATACCAAGAGAGATTAAG 192
OY 4705 TGCCTGGAGAACCAACCTCTACGCAATGAAACATGCGGGTGTGACCGCATGCT 4764
DB 193 TGTGTTGAGAAACCAACCTCTACGCAATGAAACATGCGGGTGTGACCGCATGCT 252
OY 4765 AATGACCGAGAGAGAGAGAGTGTCTCTGTAACATGCAAGAAATACATGCAAGTACTAAG 4824
DB 253 AATGACCGAGAGAGAGAGTGTCTCTGTAACGGAAGAAATATCTGCAATGCTACTAAG 312
OY 4825 CCGGACTCTATCCACTCTTCGACGGGATTTTCTCCAGCTCTAAT 4873
DB 313 CCAGACTCTTACCTTTGTCATGGAATCTTCTTCTCTCTACT 361

RESULT 30
AAD22461
ID AAD22461 standard; DNA; 3147 BP.
XX
XX AAD22461;
XX
XX 12-FEB-2002 (first entry)
XX
XX Human C3d3 DNA-pfMSPI.19 coding sequence from pVK104-02 vector.
DE Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;
XX Immune response; pVK104-02 vector; ds.
KM
XX

OS Chimeric - Homo sapiens.
OS Chimeric - Plasmodium falciparum.
OS Chimeric - Unidentified.
XX
XX Key Location/Qualifiers
FT CDS 1..3147
FT /tag= a
FT /product= "Human C3d3-pfMSPI.19 protein"
FT /note= "CDS does not include stop codon"
FT /partial
FT sig_peptide 1..72
FT /tag= b
FT mat_peptide 73..3147
FT /tag= c
FT /product= "Mature human C3d3-pfMSPI.19 protein"
FT 2845..3142
FT /tag= d
FT /note= "Antigen sequence"
XX
XX WO200177324-A1.
XX
XX 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-GB01599.
XX
XX 08-APR-2000; 2000GB-0008582.
XX
XX (ADPR-) ADPROTECH LTD.
XX
XX Steward M, Cox VF;
XX
XX WPI: 2002-010909/01.
XX
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein -
XX
XX Claim 16; Page 69-71; 87pp; English.

CC The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSPI.19-human
CC C3d3 DNA coding sequence from vector pVK104-02.
XX
XX Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other;

Query Match 3.9%; Score 194.6; DB 24; Length 3147;
Best Local Similarity 79.6%; Pred. No. 9.9e-37;
Matches 230; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 4585 AACATCTCCAGCACCATGCTGTAAGAAAGACGCCCCCAATATAGCGCTTTTCACG 4644
DB 2845 AACATGCCCCAACCAACCGCTTAAGAAAGACGCAATGTCCTCAACCTCCGATGTTCA 2904
OY 4645 CATCTGGAGACGCGGAGAGTGTCTCTGTAACATACCAAGAGAGATTAAG 4704
DB 2905 CATCTGGAGAGAGAGAGATGTCTCTGTAACATACCAAGAGAGAGATTAAG 2964
OY 4705 TGCCTGGAGAACCAACCTCTACGCAATGAAACATGCGGGTGTGACCGCATGCT 4764
DB 2965 TGTGTTGAGAAACCAACCTCTACGCAATGAAACATGCGGGTGTGACCGCATGCT 3024
OY 4765 AATGACCGAGAGAGAGAGTGTCTCTGTAACGGAAGAAATATCTGCAATGCTACTAAG 4824

```
DB 3025 AACTGCACCGAAGAACTCTGTTCTACGGAAGATTACTCGAATCTACTAAG 3084
QY 4825 CCGACTCCATACACCTTCGACGGATTTTTCGCCGCTAAT 4873
DB 3085 CCAGACTCTTACCCTTTGTTGATGGAATCTCTGTTCTCCCTACT 3133

RESULT 31
AAD22462
ID AAD22462 standard; DNA; 309 BP.
XX
AC AAD22462;
XX
DT 12-FEB-2002 (first entry)
XX
DE Plasmodium falciparum MSP1.19 (PfMSP1.19) mutant DNA.
XX
KW Immunostimulant; vaccine; immunisation; therapeutic; immune response;
XX MSP1.19 gene; mutant; ds.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT misc_feature 40..42
FT /tag= a
FT /note= "Mutation occurs at this positions which
FT results in an amino acid change from Cys to Ile"
FT misc_feature 88..90
FT /tag= b
FT /note= "Mutation occurs at this positions which
FT results in an amino acid change from Cys to Trp"
XX
PN WO200177324-A1.
XX
PD 18-OCT-2001.
XX
PF 09-APR-2001; 2001WO-GB01599.
XX
PR 08-APR-2000; 2000GB-0008582.
XX
PA (ADPR-) ADPROTECH LTD.
XX
PI Steward M, Cox VF;
XX
DR WPI; 2002-010909/01.
XX
PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein -
XX
PS Example 13; Page 72; 87pp; English.
XX
CC The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein, such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSP1.19 mutant
CC DNA used in the exemplification of the invention.
XX
SO Sequence 309 BP; 95 A; 71 C; 69 G; 74 T; 0 other;
```

Query Match 3.9%; Score 191.4; DB 24; Length 309;
Best Local Similarity 78.9%; Pred. No. 1.8e-36;

```
Matches 228; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 4585 AACATCTCCAGCAGCAATGCTGAAGAAACAGTGCCTCCAGATAGCGGCTTTACG 4644
DB 7 AACATTTGCCCAACACCAATGCTTAAAGAACAAATTCACAAACATCCGATTTTCAG 66
QY 4645 CATCTGGAGAGCGGCGAGAGAGTGTCTCTGAACTACACAAACAGAGAGATTAAG 4704
DB 67 CATCTGAGAGAGAGAGAGATGGAAGTGTCTGAACTACAAAGAGAGATGATTAAG 126
QY 4705 TGGGTGAGAGAACCAACCTTACCTGCATGAAGAAATGGCGGTGTGACCGCATGCT 4764
DB 127 TGTGTTGAGAACCCAAACCTTACCTGTAAACGAGAACAGCGTGTGACGCTGAGCT 186
QY 4765 AATGCAACGAGAGAGACAGCGCTTAAACGGAAGAAATACATGCGATGTACTAAG 4824
DB 187 AATGCAACGAGAGACAGCTGTGTTCTAACGGAAGAAATTTCTTCCGATTTACTAAG 246
QY 4825 CCGACTCCTATCCACTCTTGACGAGGATTTTGTCCAGCTCTAAT 4873
DB 247 CCAGACTCTTACCCTTTGTTGATGGAATCTCTGTTCTCCCTACT 295

RESULT 32
AAD22463
ID AAD22463 standard; DNA; 3147 BP.
XX
AC AAD22463;
XX
DT 12-FEB-2002 (first entry)
XX
DE Mutant PfMSP1.19-human C3d3 DNA coding sequence from PVK104-03 vector.
XX
KW Human; Immunostimulant; vaccine; immunisation; therapeutic; C3d;
KW Immune response; PVK104-03 vector; ds.
XX
OS Chimeric - Homo sapiens.
XX Chimeric - Plasmodium falciparum.
XX Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..3147
FT /tag= a
FT /product= "Human C3d3-PfMSP1.19 protein"
FT /note= "CDS does not include stop codon"
FT /partial
FT sig_peptide 1..72
FT /tag= b
FT mat_peptide 73..3147
FT /tag= c
FT /product= "Mature human C3d3 protein"
FT misc_feature 73..369
FT /tag= d
FT /note= "Antigen sequence"
XX
PN WO200177324-A1.
XX
PD 18-OCT-2001.
XX
PF 09-APR-2001; 2001WO-GB01599.
XX
PR 08-APR-2000; 2000GB-0008582.
XX
PA (ADPR-) ADPROTECH LTD.
XX
PI Steward M, Cox VF;
XX
DR WPI; 2002-010909/01.
XX
PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein -
XX
```

PS Claim 16; Page 73-75; 87pp; English.
XX
XX The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum mutant 'PfmSpl.19-human C3d3 DNA coding sequence from vector pVK104-03'.
XX
SQ Sequence 3147 BP; 870 A; 754 C; 834 G; 689 T; 0 other;

Query Match 3.9%; Score 191.4; DB 24; Length 3147;
Best Local Similarity 78.9%; Pred. No. 6e-36;
Matches 228; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 4585 AACATCTCCAGCACCAATGCGTGAAGAAAGAGTCCCGCAATAGCGGCTTTTCAGG 4644
DB 73 AACATTTGCCCAACCAATGCGTTAAGAAACAAATTCACAAACTCCGATGTTTCAGA 132
OY 4645 CATCTGGACGAGCGGAGTGCAGAGTCTCTGACTCAACACAGAGAGATAG 4704
DB 133 CATCTGGACGAGAGAGAAATGAAATGCTGTGAACTACACAGAGAGATATAG 192
OY 4705 TGCCTGAGAACCCAAACCTACCTGCAATGAAACAAATGCGGCTGAGCCGATCT 4764
DB 193 TGTGTGAGAACCCAAACCTACCTGCAATGAAACAAATGCGGCTGAGCGCTACCT 252
OY 4765 AAATGACACCGAGAGACGCGCTCTACGAGAAAGAAATACATGCGAGTACTAAG 4824
DB 253 AAGTGCACCGAAGAGACTGCTTAACGGAAGAAAGATTAATGCGAATGTTACTAAG 312
OY 4825 CCCGACTCTTACCTTGTGAGCGGATTTTGTCTCCAGCTCTAAT 4873
DB 313 CCAGACTCTTACCTTGTGAGCGGATTTTGTCTCTCCCTTACT 361

RESULT 33

AAD22464
ID AAD22464 standard; DNA: 3147 BP.

XX AAD22464;

DT 12-FEB-2002 (first entry)

XX Human C3d3-PfMSP1.19 mutant DNA coding sequence from pVK104-04 vector.

XX Human; immunostimulant; vaccine; immunisation; therapeutic; C3d3;

KW Immune response; pVK104-04 vector; ds.

OS Chimeric - Homo sapiens.

OS Chimeric - Plasmodium falciparum.

OS Chimeric - Unidentified.

XX Key Location/Qualifiers

FT CDS 1..3147

FT sig_peptide

FT mat_peptide

FT /product= "Mature human C3d3-PfMSP1.19 protein"

PN WO200177324-A1.
XX
XX 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-GB01599.
XX
XX 08-APR-2000; 2000GB-0008582.
XX
XX (ADPR-) ADPROTECH LTD.
XX
XX Steward M, Cox VF;
XX
XX WPI; 2002-010909/01.
XX
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
XX occurring protein and comprises a sequence non-identical to naturally
XX occurring DNA sequence encoding the protein

Claim 16; Page 76-78; 87pp; English.

CC The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfmSpl.19-human
CC C3d3 DNA coding sequence from vector pVK104-04.
XX
SQ Sequence 3147 BP; 870 A; 754 C; 834 G; 689 T; 0 other;

Query Match 3.9%; Score 191.4; DB 24; Length 3147;
Best Local Similarity 78.9%; Pred. No. 6e-36;
Matches 228; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 4585 AACATCTCCAGCACCAATGCGTGAAGAAAGAGTCCCGCAATAGCGGCTTTTCAGG 4644
DB 2845 AACATTTGCCCAACCAATGCGTTAAGAAACAAATTCACAAACTCCGATGTTTCAGA 2904
OY 4645 CATCTGGACGAGCGGAGAGTGCAGAGTCTCTGACTCAACACAGAGAGATAG 4704
DB 2905 CATCTGGACGAGAGAGAGAAATGAAATGCTGTGAACTACACAGAGAGAGATAG 2964
OY 4705 TGCCTGAGAACCCAAACCTACCTGCAATGAAACAAATGCGGCTGAGCCGATCT 4764
DB 2965 TGTGTGAGAACCCAAACCTACCTGTAACGAGAAACAGGATGAGCGCTACCT 3024
OY 4765 AAATGACACCGAGAGACGCGCTCTAACGAGAAAGAAATACATGCGAGTACTAAG 4824
DB 3025 AAGTGCACCGAAGAGACTGCTTAACGGAAGAAAGATTAATGCGAATGTTACTAAG 3084
OY 4825 CCCGACTCTTACCTTGTGAGCGGATTTTGTCTCCAGCTCTAAT 4873
DB 3085 CCAGACTCTTACCTTGTGAGCGGATTTTGTCTCTCCCTTACT 3133

RESULT 34

AAN50354
ID AAN50354 standard; DNA: 306 BP.

XX AAN50354;

DT 22-OCT-1991 (first entry)

XX Storage-specific, late schizont merozoite malaria antigen insert of

DE plasmid p31-1.

```
KW Malaria; vaccine; plasmid p31-1; ss.
XX
XX Plasmodium spp.
OS
XX
XX
FH Key Location/Qualifiers
FT mat_peptide 3..306
FT /*tag= a
FT /label= malaria antigen
XX
XX
XX W08503725-A.
XX
XX
XX 29-AUG-1985.
XX
XX 20-FEB-1985; 85MO-GB00072.
XX
XX 20-FEB-1984; 84GB-0004378.
XX
XX (BIO ) BIOGEN NV.
XX
XX Mach B, Perrin L, McGarvey M, Cheung A, Shaw A;
XX
XX WPI: 1985-223371/36.
XX
XX P-PSDB; AAR50303.
XX
XX Prodn. of antigens of Plasmodium species, esp. of falciparum - by
XX PT recombinant DNA methods giving polypeptide(s) for protecting
XX PT against malaria or for diagnosis.
XX
XX PS Disclosure: Fig 5; 49pp; English.
XX
XX CC The DNA sequence encodes a Plasmodium falciparum, Plasmodium vivax,
XX CC Plasmodium malariae and Plasmodium ovale antigen, which may be used
XX CC in the diagnosis of malaria and as a vaccine against malaria.
XX CC Nucleotides 293-422 and 429-530 are used in the preparation of
XX CC monoclonal antibodies against the malaria antigen.
XX
SQ Sequence 306 BP; 87 A; 49 C; 82 G; 88 T; 0 other:
Query Match 2.4%; Score 119.2; DB 6; Length 306;
Best Local Similarity 70.4%; Pred. No. 6; 8e-19;
Matches 159; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 82 TATCAGAGCGTGGTTAAGAACTGGAAGCTTTGAGAGTCCCTTACCGGATACAGC 141
DB 18 TATCAAGAACTGTCTCAAAAACTAGAGCTTTAGAGATCCTATTGACAGGTTATAGT 77
QY 142 CTGTTCCAGAGAGAGAGATGCTGATGAAGGAGCAGCGGCGCTTACCAAC 201
DB 78 TTATTTCAAAAGAAAAAATGATTAATGAAGAAACAGTGGACAGCTGTACACT 137
QY 202 AGCAGACCCGGTTCTAAAGGCTGTGGCTAGCGGTGGCTCCGGTGGCTGTGGCCTCT 261
DB 138 AGTACACCTGTGCTCAGGTGCTTCACTTCAAGTGTTCAGGTGGTTCAAGTTGCTTCA 197
QY 262 GGGGGTTCGTCGCTCCGCGGCGCAGCGTGCATCAGAGTGGCTCAG 307
DB 198 GTTGTTCAGGTGCTTCAGGTGCTCAGTTCAGTTCAGTTCAGTTCAG 243
RESULT 35
AAT05868
ID AAT05868 standard; DNA; 3399 BP.
XX
XX AAT05868;
XX
XX
XX 14-AUG-1996 (first entry)
XX
XX Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
XX
XX Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
XX immunisation; vaccination; ss.
XX
XX Chicken leucocytozoan.
```

```
XX
XX Key Location/Qualifiers
FH CDS 1..3399
FT /*tag= a
FT misc-feature 1150..3218
FT /*tag= b
FT /note= "Fragment referred to in the claims, for
FT use as insert in a recombinant vaccine
FT against chicken leucocytozoan disease"
XX
XX
XX JP07284392-A.
XX
XX
XX 31-OCT-1995.
XX
XX 19-APR-1994; 94JP-0080643.
XX
XX 19-APR-1994; 94JP-0080643.
XX
XX (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
XX PA (KITA ) KITASATO KENKUSHO SH.
XX
XX WPI: 1996-006311/01.
XX
XX P-PSDB; AAR97866.
XX
XX Chicken leucocytozoan immunogenic protein - used in a recombinant
XX PT vaccine against chicken leucocytozoan disease
XX
XX PS Claim 6; Page 6-9; 35pp; Japanese.
XX
XX AAT05868 encodes a chicken leucocytozoan immunogenic protein, this DNA
XX CC or a fragment of it can be used in a recombinant vaccine to immunise
XX CC against chicken leucocytozoan disease. The DNA is used in a vector
XX CC and operatively linked to an expression regulatory sequence as in
XX CC standard practice.
XX
SQ Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other:
Query Match 2.8%; Score 105.6; DB 17; Length 3399;
Best Local Similarity 43.8%; Pred. No. 4; 8e-15;
Matches 563; Conservative 0; Mismatches 709; Indels 12; Gaps 2;
QY 702 CAAGGACAAATGTGGAAAGTGGAGATTTATTTAAAAAGATAAGAGACATCGAA 761
DB 1798 CATGAGAGAGAAAAAGAGAGATGACATGAGAGAAAAAGAGAGCATGAGAA 1857
QY 762 CATTAGAGAGCTGATGAGAAATCCAAAAAGACATGACAAATTAAGATCAACCA 821
DB 1858 GTATATCATGAGAGAGAAAAAGAGAGTAACATGAGAGAAATTAAGAAAGAGCAT 1917
QY 822 GGAAGAGAGAAAAAGAGAGTTGTACAGGCCCGCTGACAGCTCTCTATTAACAACA 881
DB 1918 GAAAGAGTAAATACATGAGAGAAAAAGAGAGTAACATGAGAGAAATTAAGAAAGAA 1977
QY 882 GCTTGAAGAGCCCATTAACCTCATCAGCTGAGAGAGAGCGCATAGACACCTCAAGAA 941
DB 1978 GAGCATGAGAGAGTAATACATGAGAGAGAAAAAGAGAGTAACATGAGAAATTAAGAA 2037
QY 942 GAATGAGAGATATCAAGAGTGTCTCGACAGATTAATGAATTAAGATTCCTCCGCGAGC 1001
DB 2038 AAGAGAGAGCATGAGAGAGTAATACATGAGAGAGAAAAAGAGAGTAACATGAGAGAA 2097
QY 1002 CAACCTTGGAACACCCCTTACACGCTGTGACAGAGAGAGAGAGATAGAGAGACCA 1061
DB 2098 ATAGAAAAAGAGAGCATGAGAGAGTAATACATGAGAGAGAAAAAGAGAGTAACAT 2157
QY 1062 GAAAGAGATCAAGAGATGCGCCCAAAACCATTAAGTCAACATGATTCCTCTTTACTGA 1121
DB 2158 GAAAGAGAGAGAGAGAGAGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGTA 2217
QY 1122 TCCCTTGAGCTGAGTACTTGTGAGAGAGAGAGATTAAGATTAAGACATCTCCGCCAA 1181
DB 2218 ACACATGAGAGAGAGAGAGAGAGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGTA 2277
```


OY	1182	AGTCGACGACGAAGAACCAACCCGAACCTAATGAAATATCCCAATGGGTGACGACCTCT	1241
Db	2278	GAGAGAGAAAAAGAAAGTAATATCATGAAAGAAAGAAAAAGAA-----GAAGTAAACACAT	2331
OY	1242	GTCCTTAATACGATATCAACAACGCTCTCAACGAGCTCAATAGCTCGGTGACTTGTATTA	1301
Db	2332	GAGAGAGAAAAAGAAAGTAATACATGAAAGAAAGAAAAAGAAAGTAATACACATGAAGAA	2391
OY	1302	CCCCCTGCATTTATACGAAAGAACCCCTTAAGATATCTACACGACAAATGAGAGAAAGAA	1361
Db	2392	GA-----AAAAGAGAAAGTAATACATGAAAGAAAGAAAAAGAAAGTAATACACATGAAGAA	2445
OY	1362	GTTTATCAACGAATTCAAAGGACGATCAAAATTTGAGAGAAAGAAATTTGAGGTGACAA	1421
Db	2446	GAGAGAAAAAGTAACACATGAAAGAAAGAAAAAGAAAGTAATACATGAAAGAGAAAAAGAA	2505
OY	1422	GAAAGTTACGAAAGACCGCAGCAAAAGCTTTAAACGATATCTATAAGAGTATGAAAAAGCT	1481
Db	2506	GAAATTAACCATGAAAGAAAGAAAAAGTAATACATGAAAGAAAGAAAAAGTAATACAT	2565
OY	1482	GCTGAACAGAGATCTATGATTCGCAATTCACCAATTAACATGACCTGACCACTGAGAGAA	1541
Db	2566	GAGAGAGAGAAAAAGTAACATGAAAGAAAGAAAGAAAGAAAGAAAGAAAAAGTAACACAT	2625
OY	1542	AATGATGGGAAACGGTACTTTACAAAGGTGAGAAATCGACACACATTAATACCTTTGC	1601
Db	2626	GAGAGAGAGAAAAAGTAACACATGAAAGAAAGAAAAAGTAATACACATGAAAGAAAGAA	2685
OY	1602	ATCTATATAGAAATTTAAGCATTAATCTTTGAGAAAGCTCACCAGTAAGCTTTAATATGGA	1661
Db	2686	AAAGTAATACATGAAAGAAAGAAAAAGAAAGATGAGGAAAGAAAGAAAGAAAGAA	2745
OY	1662	GGACATATCTCTGGCGAACATTTGTTGGAGAAAGAACTTAAGTATTTACAGATCTCAT	1721
Db	2746	GAGAGAGAGAAAGAAAGAAAGAAAGAAAGATGAGGAAAGAAAGAAAGAAAGAAAGAT	2805
OY	1722	AAATAGATCGAAAAACGATCGAGCGCTTTGAGAACATTAAAGAGATGAAGAGACA	1781
Db	2806	GAGGAGAGGAAAGAAAGAAAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGTAAGAGAA	2865
OY	1782	GTTTGTGAGAGAGAAATTCGAAAGCGAAATTAACGAGATGAGAGATCTCGAGGT	1841
Db	2866	GAGAGAGAGAAAAAGAAAGCATGAGAGAAAGTAACCATGAAAGAGAAAGAGAAAAA	2925
OY	1842	CTCCGATATTTGTTAAAGTCCAGGTGCAGAAAGGTGCTCTCTCATGAACAGATTGATGAAT	1901
Db	2926	GTAACACATGAAGAAAGAAAGAAAGTAACATGAAAGAAAGAAAGAAAGTAATACATATGAA	2985
OY	1902	CAGAGAGCTAATCACTTCTTGAAGAACGTGGAGTTTAAACATATATATCATGTGCCGA	1961
Db	2986	GAGAGAGAGAAAAAGTAACACATGAAAGAAAGAAAGAAAGTAACATGAAAGAGAA	3045
OY	1962	TAGTTATTAAGCAGAGAGAAATAGCA	1985
Db	3046	AAAGTAACACATGAAAGAAAGAA	3069
RESULT 36			
AAS00655			
ID	AAS00655 standard; DNA; 618 BP.		
XX	AAS00655;		
AC			
DT	07-SEP-2001 (first entry)		
XX			
XX	Plasmodium vivax (first entry)		
DE	Merozoite surface protein; malaria; blood; serum; diagnosis; vaccine; ds		
KW	merozoite surface protein; malaria; blood; serum; diagnosis; vaccine; ds		
KW	antigen-antibody composite; Enzyme Linked Immunosorbent Assay; ELISA;		
MS			
XX	Plasmodium vivax.		
XX			

FH	Key	Location/Qualifiers
FT	CDS	1..618
FT		/*tag= a
FT		/partial
FT		/product= "P. vivax merozoite surface protein
FT		C-terminal region"
FT		/note= "No start or stop codon"
PN		WO200136587-A2.
PD		25-MAY-2001.
XX		
PF		15-NOV-2000; 2000MO-KR01302.
XX		
PR		15-NOV-1999; 99KR-0050616.
PR		25-APR-2000; 2000KR-0022041.
XX		20-MAY-2000; 2000KR-0027305.
PA		(HUMA-) HUMANDIO CO LTD.
XX		
P1		Park H;
XX		
DR		WPI: 2001-343809/36.
DR		P-PSDB; AAU00669.
PT		New gene encoding merozoite surface protein of Plasmodium vivax, useful
PT		for producing protein for diagnosis of malaria and for vaccination -
XX		
PS		Claim 5; Page 23; 24pp; English.
XX		
CC		The sequence represents a DNA encoding the Plasmodium vivax merozoite
CC		surface protein (MSP) C-terminal region. The C-terminal region of the
CC		merozoite surface protein has a strong antigenicity in malarial diseases.
CC		For diagnosis of malaria, recombinant proteins with enhanced
CC		antigenicity, obtained by addition of fusion proteins to surface protein
CC		C-terminal regions, can be reacted with serum or blood of a Plasmodium
CC		infected patient. Antigen antibody composites will be formed, and these
CC		are detected by Enzyme Linked Immunosorbent Assay (ELISA). The
CC		recombinant antigens provide a quick and reliable diagnosis of malaria,
CC		with good sensitivity and selectivity.
XX		
SQ		Sequence 618 BP; 207 A; 114 C; 147 G; 150 T; 0 other;
	Query Match	2.1%; Score 104; DB 22; Length 618;
	Best Local Similarity	54.9%; Pred. No. 4.9e-15;
	Matches 234; Conservative	0; Mismatches 180; Indels 12; Gaps 1
OY	4488	TCTCCTGACCAAGTTTGTCTGCATCGCATGGTGTCGCCAAACCTTCGCCAAGACAGTCT 4547
Dd	189	TCTTCTGGAAAATTGATGCAAAATCAAAATTTGAATAAGAAACGAGTCCAAAGCAATATT 248
OY	4548	GAGCAATCGTCGCGAGCGCAACTGCGAGGGCATGCGAACATCTCCAGCACCAATGCGT 4607
Dd	249	ATCCCACTGCTAAATTGTCACAACCTCAGTTATTAACTATGACCTCCGACCACACATGTAT 308
OY	4608	GAGAAGACAGTGGCCCCCGAAGATAGCGGCTGTTTCAGGCGATCGAGCGAGCGAAGAGTG 4667
Dd	309	AGACACCAATGTCCTGATTAATGACAGCCTGCTATAGTACTTGAGACGGAATGAAAGATG 368
OY	4668	CAAGTGTCTCTGCAACTACAAACAAGAGAGATAGTSCGTGGAGAACCCCAAACCTTAC 4727
Dd	369	GAGATGCTGTGTTAACTTTAAAGAAAGAGCGCAAGTGTGGCCAGATCGAATGTCAGC 428
OY	4728	CTGCATGAAAAATGAGGGGGGTGGAGCGCCATGCTAAATGCACCGAGGAAGACAGCG 4787
Dd	423	TTGTAAAGATTAACAATGGTGGTGTGGCCCTTGAAGTGAATTTAAATGACGACAGC-- 486
OY	4788	CTCTAACGAAAGAAATACATGCGAGTGTACTAAGCCGCACTCTTATCAGCTTTGCA 4847
Dd	487	-----AATAAATGCTGTGTAATGTGCAAAAAGAAAGGTTTGAGCCACTCTTGA 536
OY	4848	CGGATTTTTTGTGCTCCAGCTTAATTCTCGGGGCAATCTCTTCTGCTGATCTTATGCT 4907

Db	537	GGGCGTTTCTGCTAGACCTCCACGCTCCTAAGCTGTGCTCTCTGTGTGCTGATGTTGCT	536
Oy	4908	GATCCT	4913
Db	597	TTTCTCT	602

```

RESULT 37
AAT93729
ID AAT93729 standard; DNA; 165 BP.

```

DE DNA encoding signal peptide 2 which is used in a malaria vaccine

KM Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
KM MSA-1; recombinant vaccinia virus; Plasmodium falciparum;
KM immune response; humoral; cell-mediated; merozoite; ss.

OS Mammalia.

PN WO9726911-A1.

PD 31-JUL-1997

PF 29-JAN-1997; 97WO-US01395

PR 29-JAN-1996; 96US-0593006.

PA (GEOU) UNIV GEORGETOWN.

PI Davidson EA, Yang S;

DR WPI; 1997-393372/36.

DR P-PSDB; AAW34596.

PT Malaria vaccine - comprises expression vector expressing fragment of
PT merozoite surface antigen

PS Disclosure; Page 19; 75pp; English.

CC The present sequence encodes a signal peptide that is used in a malaria
CC vaccine. A novel expression vector contains a sequence encoding an
CC immunogenic merozoite surface antigen-1 peptide (MSA-1), and a
CC mammalian signal (AAM34595-96) and/or anchoring sequence (AAM34598).

Sequence 165 BP; 65 A; 16 C; 24 G; 60 T; 0 other;

Query Match	2.0%;	Score	99.4;	DB	18;	Length	165;
Best Local Similarity	75.28;	Pred. No.	3.3e-14;				
Matches 124; Conservative	0;	Mismatches	41;			Indels	0;
						Gaps	0;

Oy 130 ACCGATACAGCCTGTCCAGAAGGAGATGGTCTGAATGAA 174
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ACAGGTTATAGTTATTTCAAAAGGAAAAAATGCGATTTAATGAA 165

RESULT	38
AAI70929	165
ID	AAI70929 standard; DNA; 16S BP.

DE DNA encoding signal peptide.

KW Merozoite surface antigen; 1; MSP-1; p15MSP-1; antigen; immunogen;
KW malaria; vaccine; Venezuela equine encephalitis virus; DVEE;
KW vector; immunisation; Plasmodium falciparum; signal peptide; ss.

OS Synthetic

PN W0200185927-A1

PD 15-NOV-2001.

PF 08-MAY-2001; 2001WO-US14716.

PR 08-MAY-2000; 2000US-202430P.

PA (GEOU) UNIV GEORGETOWN

PI Davidson E, Nikodem D;

DR WPI; 2002-049444/06.

DR P-PSDB; AAM50529.

Novel vaccine for immunizing mammals against *Plasmodium falciparum* infection, comprises a viral vector system expressing protein corresponding to specific domain of major merozoite surface protein 1 of *Plasmodium falciparum*, -

PS Disclosure; Page 20; 60pp; English.

The present sequence is that of DNA encoding a typical mammalian signal peptide sequence (see AAM50529). Such a signal peptide sequence may be incorporated into chimeric proteins of the invention that also include the Plasmodium falciparum erythrocytic stage major merozoite surface antigen (MSP-1) immunogenic peptide, p15MSP-1 (see AAM50527). The invention relates to a malaria vaccine comprising an expression vector, preferably a defective Venezuelan equine encephalitis (VEE) viral vector system, which expresses p15MSP-1 or its immunogenic fragment after administration to a patient. The p15MSP-1 protein or fragment raises a humoral and/or cell-mediated response to the erythrocytic merozoite malaria antigen, protecting the patient from a subsequent malaria infection. The VEE viral vector system continues to express antigen in the patient for a period of days, months or even years. Inclusion of a signal peptide and/or an anchor peptide sequence in the p15MSP-1 antigen produces an immunogenic response which is significantly (i.e. at least about 2 times and as much as 100 times or more) greater than the immunogenic response produced by p15MSP-1 which does not contain the signal or anchor peptide sequence.

Sequence 165 BP; 65 A; 16 C; 24 G; 60 T; 0 other;

Query Match	2.0%	Score 99.4	DB 24	Length 165
Best Local Similarity	75.28	Pred. No. 3.3e-14		
Matches 124	Conservative	0	Mismatches 41	Indels 0
				Gaps 0

PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 31099; 21bp + Sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 7588 BP; 1405 A; 1605 C; 2192 G; 2386 T; 0 other;

Query Match 1.7%; Score 85; DB 23; Length 7588;
 Best Local Similarity 44.7%; Pred. No. 7.4e-10;
 Matches 377; Conservative 0; Mismatches 460; Indels 6; Gaps 1;

QY 1247 ATATGATATCAACAAAGCTCTCTCAAGAGCTTCATAGCTTGGTGAATTAACCCCT 1306
 DB 1735 ATGAGAACACAGACCAACATTAACAGAGAACAAATCCAAAGCAAGATGAGAACCC 1676
 QY 1307 TCGATTATACGAAAGAACCCCTTAAGATATCTACAGACAATGAGAGAAAGATTTA 1366
 DB 1675 AAATATACATCCAAAGCCGAATGAGAACAAACAAATTAACATCCAAAGCAAGATC 1616
 QY 1367 TCAACGAATCAAGAGAGAGATCAAAATGAGAGAAAGAAATGAGAGACAGAAAA 1426
 DB 1615 AGAACCAACCAAGCAAGATCCAAAGATGATGATGAGAGAGCAAGCAACCAACG 1556
 QY 1427 GTTACGAAGACCGAGCAAGAGTCTAAAGCATCACTAAAGATGAGAAAGCTGTGA 1486
 DB 1555 AGAACCAACATCCAAAGCAAGATCAAAACATCGAAAGCAAGATCAGAACACCA 1496
 QY 1487 AGGAGATCTATGATTCCTCAAAATTCACAAATTAACATGACATGACCACTTGAGAAAAATGA 1546
 DB 1495 ACAATCATCCAAAGCAAGATGAGAAAGCAACCAAGCAAGATCCAAAGCAAGATGAGA 1436
 QY 1547 TGGGAAAAACGCTACTTCAAAAGTGAAGAACTGACACACCATTAATCTTGCAATCT 1606
 DB 1435 ACAACCAACCAAGCAAGATCCAAAGCAAGAAAGCAACCAACCAACCAATCCAA 1376
 QY 1607 ATGAGATTTTAAGCATATCTTGAAGAGCTCCAAAGCTCTTAATATATGAGAGACT 1666
 DB 1375 GCAAGACTCAAAACACACAGACATCTCCAAAGCAAGATGAGAACACAGAGACACT 1316
 QY 1667 ATTCTCTGCGGAACATTGTTGTGAGAGAAAGAACTAAAGTATTACAAATCTCATAGTA 1726
 DB 1315 -----TCCAAAGCAAGAAATCAGAACAAATCCAAAGCAAGAAATCAGAACACAGTACA 1262
 QY 1727 AGATCGAAAAAGAGATCGAAGCCGTTTGAAGACATTTAAGAGAGATGAAGAACAGTTGT 1786
 DB 1261 ACATTCCAACCAAGAAATCAGAACACCAAGCAAGATCCAAAGCAAGAAATCAGAAC 1202
 QY 1787 TTGGAAGAGATTTACAAAGAGAGAAATTAACCAAGATGAGAGATCTTGAGAGTCCGC 1846
 DB 1201 ACAACCAAGAACCATCCAAAGCAAGAAATCAGAACATGCAAGCAAGAAATCAGAACAC 1142

QY 1847 ATATGTTAAAGTCCAAAGTGCAGAGGTGCTCTCATGAACAAAGATTGAACTCAGAA 1906
 DB 1141 ACCAGATTAACTCAAGCAAGAAATCAAGAAACCAACCAAGAAATTAACATCCAAAGCA 1082
 QY 1907 AGACTCAACTCATTTCTAAGAGCTGAGATTAAACATTAATATATCTGCGCAATAGTT 1966
 DB 1081 ATCAGAACCAACCAAGCAAGATTAACATCCAAAGCAAGATGAGAACCAACCAAGATTA 1022
 QY 1967 ATAGCAGAGAAATTAACAGAACCAAGTACTGATCTGATCTGATCTGATCTGATCTGAT 2026
 DB 1021 TCCAAAGCAAGATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 962
 QY 2027 AACTGAAAGTGTCTCAGTCCCAAGTGCAGAGCTGATCAAGAGAGAGAGAAACATTA 2086
 DB 961 AAAGCAAGAAATCAGAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 902
 QY 2087 AAA 2089
 DB 901 AAA 899

RESULT 41

AA087587
 ID AA087587 standard; DNA; 1686 BP.
 XX
 AC AA087587;
 XX
 DT 19-DEC-1995 (first entry)
 XX
 DE DNA encoding Leucocytozoan protozoa structural protein epitope.
 XX
 KW Leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;
 XX Leucocytozoanosis; treatment; ss.
 XX
 XX Leucocytozoan protozoa sp.

OS JP07089995-A.
 PN 04-APR-1995.
 PD 10-SEP-1993; 93JP-0226078.
 PF 10-SEP-1993; 93JP-0226078.
 PR 10-SEP-1993; 93JP-0226078.
 XX
 PA (DOBU-) DOBUTSUO SEIBUTSUGAKUTOKI SEIZAI KYOKAI.
 XX (NISE-) NISEIKEN KK.
 XX
 DR WPI; 1995-167252/22.
 DR P-SDB; AAR70491.
 XX
 PT Immune inducing polypeptide against Leucocytozoan protozoa - useful
 PT in production of vaccines for treatment of leucocytozoanosis in
 PT fowl.
 XX
 PS Claim 1; Page 12-14; 20pp; Japanese.

AA087587-89 encode polypeptides having a whole or partial epitope of a
 CC structural protein of Leucocytozoan protozoa (see AAR70491-93). The
 CC polypeptides and DNA encoding them are useful in the production of
 CC vaccines for the treatment of leucocytozoanosis of fowl.
 CC
 XX

Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 other;

Query Match 1.7%; Score 81.6; DB 16; Length 1686;
 Best Local Similarity 44.4%; Pred. No. 2.3e-09;
 Matches 327; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

QY 1354 AGAAGAGAGTTTATCAAGCAAGATCAAGAGATCAAAATGAGAGAGAGAAATTTAG 1413
 DB 86 AAATGTAG 145
 QY 1414 AGTGAACAG 1473

Db 146 AACAGAACAGAGAACAGAACAGAAATCGTAGAAGACAGAACAGATGAAGAAG 205
 QY 1474 GAAAGCTCTGACAGATCTATGATTCCTCAATTCAGAAATACCTCCAGCTGACAC 1533
 Db 206 AACAGAACAGAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265
 QY 1534 TTGAGAGAAATGAG 1593
 Db 266 AAGAGAACAG 325
 QY 1594 ACCCTTGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1653
 Db 326 AAGAGAACAG 385
 QY 1654 TATGAG 1713
 Db 386 AATCTCATAG 445
 QY 1714 AATCTCATAG 1773
 Db 446 AAGAGAACAG 505
 QY 1774 GAGAGACAG 1833
 Db 506 AAGATGAG 565
 QY 1834 CTGAGAGCTCGATATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1893
 Db 566 ATGCTGAG 625
 QY 1894 GATGAG 1953
 Db 626 AAAAAATATATGAG 685
 QY 1954 GTGCGGAG 2013
 Db 686 ATTTTAAAG 745
 QY 2014 AAG 2073
 Db 746 ATTAG 805
 QY 2074 AAG 2089
 Db 806 AAG 821
 RESULT 42
 ABL33103/c
 ID ABL33103 standard; DNA: 7758 BP.
 XX
 AC ABL33103;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1076.
 XX
 KW Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antileukemia; antineoplastic; antineoplastic; antineoplastic;
 KW antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 KW antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 KW antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene: ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000: 2000DE-1032529.
 PR 01-SEP-2000: 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1: SEQ ID NO 1076; 32bp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 7758 BP; 1608 A; 44 C; 188 G; 4268 T; 0 other;
 Query Match 1.7%; Score 81.6; DB 24; Length 7758;
 Best Local Similarity 42.6%; Pred. No. 5,1e-09;
 Matches 481; Conservative 0; Mismatches 644; Indels 3; Gaps 1;
 QY 693 CGACACATCAAG 752
 Db 2263 CTRAAAC 2204
 QY 753 CATGAG 812
 Db 2203 CCAC 2144
 QY 813 TGCAC 872
 Db 2143 CCAC 2084
 QY 873 TAAC 932
 Db 2083 CCAC 2024
 QY 933 CCTCAAG 992
 Db 2023 TCTAAAC 1964
 QY 993 TCCGCGAC 1052
 Db 1963 CTAAAC 1904
 QY 1053 GGAGCAG 1112
 Db 1903 CTRAAAC 1844
 QY 1113 CTTTACTGATCCCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1172
 Db 1843 CCAC 1784
 QY 1173 CTCCGCGAC 1232
 Db 1783 CTAAAC 1724
 QY 1233 GTACCTCTGCTTATAGAGATATCAACAGCTCTCAAGAGAGAGAGAGAGAGAG 1292
 Db 1723 CTRAAAC 1664
 QY 1293 CTTGATTAAAC---CCCTTGATTTATAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1349

Db 1663 CCAAAACACAAATCTACAACTAAACAACTAAACAAACAAACAAACAAACAAAC 1604
QY 1350 TGAGAGAAAGAGTTTATTCACGAAATTCAGAGAGATCAAAATTCAGAGAAAT 1409
Db 1603 CTAACAAAAACAAAAACAACTAAACAACTAAATACACAAAAACAACTAAACAT 1544
QY 1410 TGAGAGTACAGAAAGATTCAGAAACCCAGCAAGATCTAAACGATATCACTAAGA 1469
Db 1543 CTAATACACAAAAACAACTAAACAAACAAACAAACAAACAAACAACTAAACAA 1484
QY 1470 GTATGAAACAGCTGTGACAGATCTATGATTCCAATTCACATATCACTGACCTGAC 1529
Db 1483 CAATATAAACAACTAAACAACTAAACAAACAAACAAACAACTAAACAA 1424
QY 1530 CAACCTTCAGAAAAATGATGGAAAAACGCTCTTACAAAGTGAAGAACTGACACCA 1589
Db 1423 CAAACAAACAACTAAACAACTAAACAAACAACTAAACAAACAACTAAACAA 1364
QY 1590 TAATACCTTTGCAATCTATGAGAAATTCATGATGAGAACTGACCAAGCTCT 1649
Db 1363 AAAAAAAACAACTAAACAACTGCAAAACAAACAAACAAACAAACAACTAAACAA 1304
QY 1650 TAAGTATATGAGAGACTATCTCTGCGGAACATTTGTGTGAGAAACAACTAAAGTATTA 1709
Db 1303 CAAAAAACACAACTAAACCTCCAAACAAACAAATTAACAAACAACTAAACAA 1244
QY 1710 CAAGAATCTCATAGTAAGATGCAAAACGAGATGAGAGCTGTTGAGAACTTAAGA 1769
Db 1243 CAAAAAAACAACTAAACAACTGCAAAACAAACAAACAAACAACTAAACAA 1184
QY 1770 GGATGAAACAGTTGTTGAGAAAGATTAACAAAGACGAAATTA 1817
Db 1183 CAAAAAAACAACTAAACAACTGCAAAACAAACAAACAAACAACTAAACAA 1136

RESULT 43
AAH47054 standard; DNA: 5438 BP.

AC AAH47054;
XX
XX 29-OCT-2001 (first entry)
DE Synthetic gene ViVac encoding recombinant protein ViVac1p.
XX
XX Multivalent protein; Immune response; Plasmodium vivax; parasite;
KM protozooid; vaccine; malaria; recombinant; ViVac1; ds.
XX
OS Synthetic.
OS Plasmodium vivax.
XX
XX
FH Key Location/Qualifiers
FT CDS 7..5430
FT /tag= a
FT /product= "ViVac1p"
XX
XX
PN WO200155181-A2.
PD -02-AUG-2001.
XX
XX 29-JAN-2001; 2001MO-US02937.
XX
XX 31-JAN-2000; 2000US-0179213.
PR
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX Lai AA, Xiao L, Zhou Z;
XX
XX WPI: 2001-514557/56.
DR P-PSDB: AAB85697.
XX
XX
PT New recombinant multivalent protein comprising antigenic determinants
PT derived from more than one stage in a life cycle of Plasmodium vivax.

PT useful as a vaccine for treating, preventing and reducing malaria
PT infection
XX
XX Claim 9; Page 36-39; 59pp; English.
PS
XX
XX The invention relates to recombinant multivalent proteins (I) that
CC stimulate an immune response to Plasmodium vivax. (I) comprises antigenic
CC determinants, fragments or conservative substitutions, derived from more
CC than one stage in a life cycle of a Plasmodium vivax parasite. (I) is
CC useful as a vaccine for stimulating an immune response, specifically a
CC protective immune response that confers increased resistance to infection
CC by Plasmodium parasites, such as P. vivax. (I) is especially useful in
CC the treatment, prevention and reduction of malarial infection, as
CC research or diagnostic reagents for the detection of Plasmodium species
CC in a biological sample, and for conferring immunity against multiple
CC stages of the malarial parasite. The antibodies produced are useful for
CC the detection or measurement of antigenic epitopes derived from one or
CC more stages in a life cycle of a parasite, particularly P. vivax. The
CC vaccine comprising the recombinant proteins, is cost-effective, health-
CC promoting intervention for controlling, preventing or treating the
CC incidence of malaria. The present sequence represents a synthetic gene
CC ViVac1 encoding the recombinant protein ViVac1p, a multivalent and
CC multistage vaccine against P. vivax.
XX
XX
SQ Sequence 5438 BP; 1861 A; 1038 C; 1389 G; 1150 T; 0 other;
Query Match 1.5%; Score 72.4; DB 22; Length 5438;
Best Local Similarity 55.0%; Pred. No. 7.3e-07;
Matches 142; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 4588 ATCTCCAGACCAATGCGGAAGAAACAGTGCCTGAGATGAGCGCTTTTCAGCAT 4647
Db 5164 AGCTCCGAGACATGTAAGACCAATGCTGATGATGAGCGCTCTTATAGGTAC 5223
QY 4648 CTGAGACGAGCGCGAGAGTGCAGTGTCTGACTACAAACAAAGAGATTAAGTGC 4707
Db 5224 TTGGACGGAACGAGATGAGATGCTTTAAAGAAAGAGCGCAAGTGT 5283
QY 4708 GTGAGAGAACCCCAACCTTACCTGCAATGAACAAATGCGGCTGAGCGCATGCTAA 4767
Db 5284 GTGCCACATCGAATGATGCTTTAAAGATTAAGATGATGCTGCGCTGAAGCTGAA 5343
QY 4768 TGCACGAGAGACAGCGGCTCTACGGAAGAAATCAATGCGATGACTAAGCCC 4827
Db 5344 TGTAAATGACGACCAATTAATGCTGTAAATGACTAAAGAGGTTTCAGCCA 5403
QY 4828 GACTCTATTCACATCTTC 4845
Db 5404 CCTTTGAGGAGTTTTC 5421

RESULT 44
AAH47055 standard; DNA: 6101 BP.

AC AAH47055;
XX
XX 29-OCT-2001 (first entry)
DE Synthetic gene ViVac2 encoding recombinant protein ViVac2p.
XX
XX Multivalent protein; Immune response; Plasmodium vivax; parasite;
KM protozooid; vaccine; malaria; recombinant; ViVac2; ds.
XX
XX
OS Synthetic.
OS Plasmodium vivax.
XX
XX
FH Key Location/Qualifiers
FT CDS 7..6093
FT /tag= a
FT /product= "ViVac2p"
XX
XX
PN WO200155181-A2.

```
XX 02-AUG-2001.
PD
XX
XX 29-JAN-2001; 2001WO-US02937.
PF
XX 31-JAN-2000; 2000US-0179213.
PR
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX Lal AA, Xiao L, Zhou Z;
PI
XX WPI; 2001-514557/56.
DR
XX P-PSDB; AAB85698.
PT New recombinant multivalent protein comprising antigenic determinants
PT derived from more than one stage in a life cycle of Plasmodium vivax,
PT useful as a vaccine for treating, preventing and reducing malarial
PT infection -
XX
XX Claim 9; Page 45-48; 59pp; English.
XX
XX The invention relates to recombinant multivalent proteins (I) that
CC stimulate an immune response to Plasmodium vivax. (I) comprises antigenic
CC determinants, fragments or conservative substitutions, derived from more
CC than one stage in a life cycle of a Plasmodium vivax parasite. (I) is
CC useful as a vaccine for stimulating an immune response, specifically a
CC protective immune response that confers increased resistance to infection
CC by Plasmodium parasites, such as P. vivax. (I) is especially useful in
CC the treatment, prevention and reduction of malarial infection, as
CC research or diagnostic reagents for the detection of Plasmodium species
CC in a biological sample, and for conferring immunity against multiple
CC stages of the malarial parasite. The antibodies produced are useful for
CC the detection or measurement of antigenic epitopes derived from one or
CC more stages in a life cycle of a parasite, particularly P. vivax. The
CC vaccine comprising the recombinant proteins, is cost-effective, health-
CC promoting intervention for controlling, preventing or treating the
CC incidence of malaria. The present sequence represents a synthetic gene
CC vlvac2 encoding the recombinant protein vlvac2p, a multivalent and
CC multistage vaccine against P. vivax.
XX
XX Sequence 6101 BP; 2077 A; 1168 C; 1534 G; 1322 T; 0 other:
XX
XX Query Match 1.5%; Score 72.4; DB 22; Length 6101;
XX Best Local Similarity 55.0%; Pred. No. 7.8e-07;
XX Matches 142; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
XX
XX 4588 ATCTCCAGACCCATGGGTGAGAGAAACAGTCCGCCAGATAGCGGCTTTACAGCAT 4647
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 5827 AGCTCCGAGACACATGTAATGACACCAATGTGCTGATTAATGACAGCTGTATAGTAC 5886
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 4648 CTGACGAGCGGAGAGAGTGCATGCTCTGACTACAAACAGAGAGAGATAGTGC 4707
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 5887 TTGGACGAGACGGAAGATGGAGATGCTTTTAACTTTAAAGAAAGAGCGGCAAGTGT 5946
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 4708 GTGAGAAACCCAAACCTACTGCAATGAAAAAATGGCGGGGTGACGCCGATCTAA 4767
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 5947 GTGCAGCATGCAATGTGATGTAAGATACATGATGGTGTGGCCCTGAAGCTGAA 6006
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 4768 TGCACCGAGGAAGACAGCGGCTTACGGAAGAAATCAGTCGAGTGTACTAGCCC 4827
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 6007 TGTAAATATGACGAGACGAATTAATCGTGTGTAATGTACTAAAGGTTCTGAGCCA 6066
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 4828 GACTCCTATCCACTCTTC 4845
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 6067 CCCTTTGAGGAGTTTTC 6084
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX RESULT 45
XX ID AAA70099 standard; DNA: 3579 BP.
XX AC AAA70099;
XX
```

```
DT 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:232.
DE
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
XX Plasmodium falciparum.
OS
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOEFF) HOFFMAN S.
XX (CARU) CARUCCI D.
XX (GARD) GARDNER M.
XX (VENT) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
XX Disclosure: Page 457-458; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasite lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX
XX SO
XX
XX Sequence 3579 BP; 1904 A; 398 C; 552 G; 725 T; 0 other:
XX
XX Query Match 1.4%; Score 70.6; DB 21; Length 3579;
XX Best Local Similarity 47.7%; Pred. No. 1.6e-06;
XX Matches 239; Conservative 0; Mismatches 259; Indels 3; Gaps 1;
XX
XX 1610 AGAATTTAAGCATATATCTTGAAGACTCACCAAGCTTTATATATAGAGCATATT 1669
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1245 AGAAAAACAAGAAAGAAAGAGTAGAGCAAAAAAACAAGAAAAACCGAAGAAATV 1304
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1670 CTCTCGGAGACATTTGTTGGGAGAAAGAACTTAATATTACAAAGATCTCATTAAGTAAGA 1729
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1305 AGAAGAGACAAAGAAAGAAATGCAAAAAAGACAAAGCAATTCAGAAAAAGATTAAGAA 1364
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1730 TCGAAGACGAGATCGAGACGCTGTTGAGACATTTAAGAGATGAAGAACAGTTGTTTG 1789
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1365 ATCAGAAAAAGATTAAGAAATGCAAAAAAGACAAAGAAAAAAGTGAAGAGATGAAGA 1424
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1790 AGAAGAGATTTCAAAAAGCAAA---AATTAACGAGATGAGAGATCTCGAGAGTCTCGG 1846
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```


Db 1425 AAAAAGTGAAGAGCAAAAAGCAAGCAAGTATACAAAAGAAAAGACAGATGTAGATGAAA 1484
QY 1847 ATATTGTTAAAGTCCAGAGTCAGAGAGTCTCTCATGAGCAACAGATTGATGAACTCAGA 1906
Db 1485 AAAAGAAAAGGGAGAAATATGAGAGGAGCAAGATGATGAGAAAGCAAAAAGAAAAGAGA 1544
QY 1907 AGACTCAACATCTTGAAGAAACGTGAGTTAAACATATATATACATGTCCTCGAATGTGTT 1966
Db 1545 AGACGACGAGAGAAACAAATATAGAGAAAGAAAACGAGAAAAGAGCAAGAGGAAACAGA 1604
QY 1967 ATAGCAGAGAAATAGCAGAGAAACCATCTACTCATCTGACTCAGAGAAAGATGAGACA 2026
Db 1605 TTATGAAAGAGATACAGATGATTCAGACAAAGATGAAAGAAACAAAGTACAGAAAAGAA 1664
QY 2027 AACTGAAAGATTCATGAGCCCAAGTGCAGAGCCTGATCAACGAGAGCAAGAAAGACATTA 2086
Db 1665 AACAGAAAGAGACGAGAAAGAACTGAAAGAGACGAAAAGAAAACAGAGTAGAAAAAAA 1724
QY 2087 AAACCTGAAGCAGATCAGATA 2107
Db 1725 GAAACAGAAAAAGACGACAGA 1745

RESULT 46

ABQ46952/C
ID ABQ46952 standard; DNA: 646 BP.

AC ABQ46952;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 33543.

XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KM drug; side effect; cancer; central nervous system; cardiovascular;

KM gastrointestinal; respiratory system; single nucleotide polymorphism;

KM SNP; cell differentiation; ds.

XX Homo sapiens.

XX MO200218632-A2.

PD 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

PF 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

PR (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

PI WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA -

XX Claim 12: 56pp + Sequence listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one

CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the

CC label on the amplicon. From the ratio of labels hybridised to the two

CC classes of oligomers, the degree of methylation is calculated. The method

CC is used: (1) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory

CC systems etc., particularly by detecting mutations or single nucleotide

CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC types and for investigating cell differentiation. The method allows the

CC methylation status of many C residues to be determined simultaneously.

CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the

CC method for determining the degree of cytosine methylation described in

CC the disclosure of the invention.

XX SO Sequence 646 BP; 118 A; 46 C; 97 G; 385 T; 0 other;

QY Query Match 1.4%; Score 70.2; DB 24; Length 646;

QY Best Local Similarity 46.3%; Pred. No. 8.5e-07;

QY Matches 268; Conservative 0; Mismatches 308; Indels 3; Gaps 1;

Db 1258 AACACCCCTCCAAACGCTCAATAGCTTGCGTGACTTGAACCCCTCGATTAACG 1317

Db 627 AACATATGCAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 568

QY 1318 AAAGAACCCCTCTAAGATATCTACACAGACATGAGAAAGATTATCAACGAAATC 1377

Db 567 ACGATTAATAATTAACATTTAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 508

QY 1378 AAGGAGAGATCAAAATTTGAGAGAGAGAAATTTGAGAGTACAGAAAGTTACGAAAC 1437

Db 507 AAAAATTAACATTAATTAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 448

QY 1438 CGAGCAAAAGTCTAAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 1497

Db 447 AACATATGCAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 388

QY 1498 GATTCCAAATTCACAAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 1557

Db 387 ACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 328

QY 1558 TACTCTTCAAAAGTGGAGAAAGTACACACCATTAATACCTTGCA---TCCATATGAAAT 1614

Db 327 AACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 268

QY 1615 TCTAGCATTAATCTTGAAGAGTCCACCAAGCTCTTAAGTATATGAGAGCTATATCTCTG 1674

Db 267 AACATTAATAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 208

QY 1675 CGGAACATTTGTTGGAGAAAGAACTAAGTATTACAGAAATCTCATTAAGTACGAA 1734

Db 207 ACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 148

QY 1735 AACGAGATCGAGAGCGTCTTGGAGAAACATTAAGAGATGAGAAAGCTTGTGAGAG 1794

Db 147 AACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 88

QY 1795 AAGATTACAAAGACGAAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 1833

Db 87 AACATATGCAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACG 49

RESULT 47

ABQ46953
ID ABQ46953 standard; DNA: 646 BP.

AC ABQ46953;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 33544.

XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KM drug; side effect; cancer; central nervous system; cardiovascular;

KM gastrointestinal; respiratory system; single nucleotide polymorphism;

KM SNP; cell differentiation; ds.

XX Homo sapiens.

OS

```

XX WO200218632-A2.
PN
XX
XX
PD 07-MAR-2002.
PF 01-SEP-2001; 2001WO-EP10074.
XX
XX 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI: 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 646 BP; 385 A; 97 C; 46 G; 118 T; 0 other:
XX
Query Match 1.4%; Score 70.2; DB 24; Length 646;
Best Local Similarity 46.3%; Pred. No. 8.5e-07;
Matches 268; Conservative 0; Mismatches 308; Indels 3; Gaps 1;
XX
OY 1258 AACACGCGCTCAACGAGCTCAATAGCTTCGGTGAATTGATTAACCCCTTCGATTATACG 1317
DB 20 AACATATGCAATACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGAT 79
OY 1318 AAAGAACCCCTTACAGATTAATCTACACAGCAATGAGAGAAAGATTTCACAGCAATC 1377
DB 80 AGCAATTAACCAATTAACATTAATTAATTAACGATTAACGATTAACGATTAACGATTAAC 139
OY 1378 AAGGAGAAAGATCAAAATTCAGAAAGAAATTCAGAGTGCACAGAAAGATTTCAGAGAC 1437
DB 140 AAAATTAACCAATTAATTAATTAACGATTAACGATTAACGATTAACGATTAACGATTA 199
OY 1438 CGCAGCAAAAAGTCTAACGATATCACTTAAGATATGAAAAGCTGCTGAACGAGATCTAT 1497
DB 200 AACATATGCAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAAC 259
OY 1498 GATTCCAAATTCACATTAACATCGACCTGACCACTTCGAGAAAATGATGGAAAACGG 1557
DB 260 AGCAATTAACGATTAACCAATTAATTAATTAACGATTAACGATTAACGATTAACGATTA 319
OY 1558 TACTCTACAAAGTGGAGAAATGACACACCATTAATCTTTGCA---TCTATGGAAT 1614
DB 320 AACGATTAACCAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTA 379

```

```

OY 1615 TCTAAGCATTAATCTTGAGAGCTCACCAAGCTCTTAAGTATATGAGGACATATCTCTG 1674
DB 380 AACATCAAAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAAC 439
OY 1675 CGGACATTTGTTGTGAGAGAAAGAACTTAAGTATTACAGAAATCTCTAAGTAGATCGAA 1734
DB 440 AGCAATTAACGATTAACCAATTAATTAATTAATTAACGATTAACGATTAACGATTAAC 499
OY 1735 AACGAGATCGACAGCTGTTGTGAGAAATTAAGAGATGAAAGAACAGTGTGTTGGAAG 1794
DB 500 AACGATTAACCAATTAACGATTAACCAATTAACGATTAACGATTAACGATTAACGATTA 559
OY 1795 AAGATTCAAAAGACGAAATTAACGATTAACGATTAACGATTAACGATTAACGATTA 1833
DB 560 AACATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAAC 598
XX
XX RESULT 48
XX ABQ39490/C
XX ID ABQ39490 standard; DNA; 969 BP.
XX
XX ABQ39490;
XX
XX 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 26081.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP10074.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI: 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the

```

CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

SO Sequence 969 BP; 177 A; 54 C; 93 G; 645 T; 0 other;

Query Match 1.4%; Score 68.6; DB 24; Length 969;
 Best Local Similarity 43.6%; Pred. No. 2.6e-06;
 Matches 412; Conservative 0; Mismatches 524; Indels 9; Gaps 2;

```

OY 690 TCCGACACATCATGACATGAGGAGGAGATGATGATTTATTTAAAAAGATAGAA 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 939 TCTTACTAAATTTACTTAATTAACGAATACGAATATTAATACGAATTAATAATA 880
OY 750 GACCATCGAGACATTTAACGAGTGTGAGAGATCCAAAGACCATAGACAAATAAA 809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 879 AACGAATAAAAATATGCAATTAATAATAAACGAATAAAAATATGCAATTAATA 820
OY 810 GAATGCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 869
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 819 TAATTAACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 760
OY 870 CTATTAACAAACAGCTTGAAGAGCCCATACCTCATCAGCTACTGAGAGAGCCAT 929
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 759 AAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 700
OY 930 CACCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 699 GAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 640
OY 990 TCTCCGCGCAGCAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 639 AAAAAGCAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 580
OY 1050 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 579 -----GAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 526
OY 1110 TCTCTTACTGATCCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 525 TAAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 466
OY 1170 CATCTCCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 465 TAAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 406
OY 1230 GACGTAACCTCTGCTTATACGATATCAACAGCTCTCAAGAGAGAGAGAGAG 1289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 405 TACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 346
OY 1290 TGACTTGATTAACCCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 345 AACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 286
OY 1350 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 285 AAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 229
OY 1410 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 228 TACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 169
OY 1470 GTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 168 TAAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 109
OY 1530 CAACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 108 TAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49
OY 1590 TAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 48 AAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4

```

RESULT 49
 ID ABQ39491 standard; DNA; 969 BP.

AC ABQ39491;
 DT 12-JUL-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26082.
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 OS Homo sapiens.

PN W0200218632-A2.
 PD 07-MAR-2002.
 PE 01-SEP-2001; 2001MO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guettig D;
 DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
 for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridized to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridization to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridized to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX Sequence 969 BP; 645 A; 93 C; 54 G; 177 T; 0 other;

Query Match 1.4%; Score 68.6; DB 24; Length 969;
 Best Local Similarity 43.6%; Pred. No. 2.6e-06;
 Matches 412; Conservative 0; Mismatches 524; Indels 9; Gaps 2;

```

OY 690 TCTGACACATCATGACATGAGGAGGAGATGATGATTTATTTAAAAAGATAGAA 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 31 TCTTACTAAATTTACTTAATTAACGAATACGAATATTAATTAACGAATTAATAATA 90
OY 750 GACCATCGAGACATTTAACGAGTGTGAGAGATCCAAAGACCATAGACAAATAAA 809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 91 AACGAATAAAAATATGCAATTAATAATAAACGAATAAAAATATGCAATTAATAATA 150

```


DT	16-NOV-2000	(first entry)
XX	Neurospora crassa qde-1 gene.	
XX	Gene silencing; quelling deficient; qde-1; al-1; ds.	
XX	Neurospora crassa.	
OS	Neurospora crassa.	
XX	Key	Location/Qualifiers
FT	CDS	2447..6655
FT		/tag= a
FT		/product= qde-1
XX	MO2000050581-A2.	
XX	31-AUG-2000.	
XX	16-FEB-2000; 2000MO-IT00048.	
XX	22-FEB-1999; 99IT-RM00117.	
XX	(UYRO-) UNIV ROMA LA SAPIENZA.	
XX	Macino G, Cogoni C;	
XX	WPI: 2000-579171/54.	
XX	P-PSDB: AAB13956.	
XX	Novel polynucleotide encoding a polypeptide which has a silencing	
PT	activity and comprising a RNA-dependent RNA polymerase domain -	
XX	Claim 1; Page 31-43; 48pp; English.	
XX	The present sequence is the Neurospora crassa qde-1 gene. This gene has	
CC	silencing activity. The qde-1 gene was isolated by mutational analysis	
CC	of an al-1 transgenic strain. This strain had an albino phenotype	
CC	resulting from post-transcriptional silencing of the endogenous al-1	
CC	gene. Reversion of this phenotype indicated a mutation in a silencing	
CC	gene. The silencing gene, qde-1, could then be isolated. Modulation of	
CC	qde-1 expression may be used to inactivate genes and to silence	
CC	suppression of genes.	
XX		
SO	Sequence 8045 BP; 2089 A; 2180 C; 1950 G; 1826 T; 0 other;	
	Query Match 1.3%; Score 65.2; DB 21; Length 8045;	
	Best Local Similarity 50.3%; Pred. No. 5.1e-05;	
	Matches 160; Conservative 0; Mismatches 158; Indels 0; Gaps 0	
QY	1238 CTCCTCTTATATACGATATCAACACGCTCTCAACGAGCTCAATAGCTTCGGTGA	1297
DB	7390 CTCCTCTGATCATCATCGAAGACACACACACACACACACACACACACACACAC	7449
QY	1298 TTAAACCCCTTCGATTTATACGAAAGAACCTCTTACAGATATCTTACACAGCAATGAGAGA	1357
DB	7450 ACAAC	7509
QY	1358 AGAATTTTATCAACGAAATCAAGGAGGAGATCAAAATTTGAGAAAGAAATTTGAGAGTG	1417
DB	7510 ACAAC	7569
QY	1418 ACAAGAAAAATTACGAGAACCGACAGCAAAAGTCTAAAGCATATCACTAAAGATGAAA	1477
DB	7570 ACAAC	7629
QY	1478 AGCTGCTGACAGATCTATGATTTCCAAATTTAACAATATACATCGACCTGACCAACTTG	1537
DB	7630 ACAAC	7689
QY	1538 AGAAATGATGGGAAAC 1555	
DB	7690 AGACACACACACACACAC 7707	

[illegible]

```

OY 1437 CCGAGCAAAAGTCTAAACGATATCACTAAGAGATGAAAAAGCTGGAACGAGATCTA 1496
DB 2359 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACGA 2300
OY 1497 TGATTCAAATTCACATATACATGACCTGACCAACTTCGAGAAATGATGGGAAAAAG 1556
DB 2299 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACGAAAA 2240
OY 1557 GTACTCTTACAAAGTGGAACCTGACACCATATATACCTTTGATCTTGTGAGATTC 1616
DB 2239 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2180
OY 1617 TAACATATCTTGTGAGAACTCACCAGCTCTTAATATATGAGAGCTATTCCTCG 1676
DB 2179 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2120
OY 1677 GAACATTTGTTGGAGAAAGACTAAAGTATACAAAGATCTCAATAGTAAAGTGA 1736
DB 2119 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2060
OY 1737 CGAATCGAGACGCTTGTGAGACATTAAGAGATGAACAGCTTTGTGAGAGAA 1796
DB 2059 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2000
OY 1797 GATTACAAAAGACGAAATTAACGATGAGAGA 1831
DB 1999 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1965

```

RESULT 53

AAV20700 ID AAV20700 standard; DNA; 5163 BP.

AC AAV20700;

DT 17-AUG-1998 (first entry)

DE Cryptosporidium parvum GP900 antigen open reading frame.

KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;

KW antibody; prophylaxis; treatment; inhibition; retardation;

KW detection; diagnosis; human; ds.

OS Cryptosporidium parvum.

FH Key Location/Qualifiers

FT CDS 1..5163

FT misc_feature /product= GP900 antigen

FT 524..1270

FT /tag= b

FT /note= "region containing NINC mutations"

XX WO9806430-A1.

XX 19-FEB-1998.

XX 11-AUG-1997; 97MO-US14104.

XX 14-AUG-1996; 96US-0700651.

XX (REGC) UNIV CALIFORNIA.

XX Gut J, Leech J, Nelson RC, Petersen C;

XX WPI; 1998-159290/14.

XX P-PSDB; AAW48299.

XX Anti-Cryptosporidium antibody - used to develop products for

XX detection, diagnosis, prophylaxis or treatment of Cryptosporidium

XX Infections

PS Claim 32; Pages 60-62; 89pp; English.

XX The sequence is that encoding the GP900 antigen which may be used

XX in the production of anti-Cryptosporidium antibodies. These can be

XX used for the prophylaxis, treatment, inhibition or retardation of

XX a Cryptosporidium infection in humans or in animals such as calves.

XX They can also be used for the detection and diagnosis of related

XX infections.

XX Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 1.2%; Score 60.6; DB 19; Length 5163;
 Best Local Similarity 44.2%; Pred. No. 0.00053;
 Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

```

OY 3930 CAATGTCAAGTGGAAGCAATTCGACGCGCTTATATAGAGAGAAATTCAGAA 3989
DB 615 CAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 674
OY 3990 CGCTTTGAGAGCGACTTGATTCCTATTAAGACCTGACCTCTACTACTAGTTGCA 4049
DB 675 CAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 734
OY 4050 GGACCCATACAAAGTTCTCTCAATTAAGAGAGAGATTAATTTCTACTACAACTA 4109
DB 735 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 794
OY 4110 TATCAAGACTTCATCGACACCGATATCAATTTGCTATATGATGCTGGGATTA 4169
DB 795 CAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 854
OY 4170 GATCCTGAGCGAAATTAATCAAGTGTGACCTGACTCTATTAAGATATCAAGATTA 4229
DB 855 CCACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 914
OY 4230 GCAAGGCGAGATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4289
DB 915 AGAAACCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 974
OY 4290 AGTAGACACAAATCGACCTTCTTGATTAATTCACCTGAGAGCGCAAGTCTCAACTATAC 4349
DB 975 CCACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1034
OY 4350 TTACGAGAAAGCAATGTGGAAGTTAAATCAAGAGCTGAACCTCAAAACAAATCA 4409
DB 1035 CTACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1094
OY 4410 AGACAAAGCTGCGAGATTTCAAGAAATTAACAAATTTGCTGGAATTCGAGACTGTCTAC 4469
DB 1095 CAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1154
OY 4470 CGATTATACCAACAAATCTCC 4492
DB 1155 CAACCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1177

```

RESULT 54

AA61849 ID AA61849 standard; DNA; 5163 BP.

AC AA61849;

DT 28-OCT-2000 (first entry)

DE ORF encoding a portion of Cryptosporidium parvum NINC isolate GP900.

KW GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;

KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;

KW merozoite; diarrhoea; protozoa; open reading frame; ORF; ds.

OS Cryptosporidium parvum.

FH Key Location/Qualifiers

```
FT CDS 1669..7182
FT /*tag= a
FT /partial
FT /product= "Cryptosporidium parvum NINC isolate GP900"
FT /note= "No start or stop codons given in the
FT specification"
XX
XX US6071518-A.
XX
XX 06-JUN-2000.
XX
XX 12-SEP-1997; 97US-0928361.
XX
XX 13-SEP-1996; 96US-0026062.
XX 01-JUN-1993; 93US-0071880.
XX 29-MAY-1992; 92US-0891301.
XX 03-APR-1995; 95US-0415751.
XX 14-NOV-1996; 96US-0700651.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Petersen C;
XX
XX WPI: 2000-422065/36.
XX P-PSDB: AAB11727.
XX
XX New GP900 protein fragments and fusion proteins of Cryptosporidium
XX parvum, useful for detecting the presence of the parasite, and
XX diagnosing or treating Cryptosporidium infections by competitive
XX inhibition of the function of GP900 -
XX
XX Claim 16; Column 47-52; 59pp; English.
XX
XX The invention relates to the GP900 glycoprotein of the protozoan
XX Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
XX proteins comprising GP900 or fragments thereof to a host to elicit anti-
XX administration of GP900 or fragments thereof to a host to elicit anti-
XX GP900 antibody production, and to a method of cryptosporidiosis treatment
XX or prophylaxis comprising administration of anti-GP900 antibodies to an
XX individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
XX competitively inhibit sporozoite or merozoite attachment or invasion, and
XX are also useful for the generation of anti-GP900 antibodies. The
XX antibodies also inhibit sporozoite or merozoite attachment/invasion and
XX additionally inhibit the binding of GP900 ligands to GP900. GP900
XX proteins, fragments and antibodies may therefore be used to treat or
XX prevent cryptosporidiosis. Infection with Cryptosporidium is a common
XX cause of diarrhoea in humans and causes life-threatening diarrhoea in
XX immunocompromised persons. Cryptosporidiosis can be contracted from
XX contaminated municipal water supplies (e.g., public swimming pools). It
XX is also a cause of disease in animals, resulting in financial losses in
XX agriculture. GP900 fragments, fusion proteins and antibodies may also be
XX used for the diagnosis of Cryptosporidium parvum infections, and for the
XX detection of the parasite in the environment. The present sequence
XX represents the open reading frame (ORF) encoding a portion of the GP900
XX protein of the NINC isolate of Cryptosporidium parvum.
XX
XX Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 other;
XX
XX Query Match 1.2%; Score 60.6; DB 21; Length 5163;
XX Best Local Similarity 44.2%; Pred. No. 0.00053;
XX Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;
XX
XX 3930 CAATGTCACGTGAGAGACATTCTGAACAGCCGCTTTAATAGAGAAATTTCAAGAA 3989
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 614 CAACACACACACACACACACACACACACACACACACACACACACACACACACAC 673
XX
XX 3990 CGTCTTGAGAGAGCGATTGATTCCTTAATAGACCTGACCTCTTAACCTGTGCA 4049
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 674 CAACACACACACACACACACACACACACACACACACACACACACACACACACAA 733
XX
XX 4050 GAGCCGATACAAATGCTCTCAATAAGAGAGAGGATTAATTTCTGCTGTACAACTA 4109
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 734 CGACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793
```

```
QY 4110 TATCAAGAGACTCCATGACAGCCGATATCAATTTGCTAATGATGCTGGGTATTACAA 4169
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 794 CAACACACACTACTACACACACACACACACACACACACACACACACACACACAA 853
XX
QY 4170 GATTCGTGAGGAAATATACAGTCTGACCTTATTAATAAGATATTCACAGTAA 4229
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 854 CCACAACTACTACACACACACACACACACACACACACACACACACACACACAC 913
XX
QY 4230 GCAAGGGAGATGAAATATATCTGCCCTTCGAAATACATGAAACCTGTACAGAG 4289
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 914 AGAACCCACACACACACACACACACACACACACACACACACACACACACACAC 973
XX
QY 4290 AGTGAACGACAAATATGACCTCTTCTTAATTCACCTGAGGCCAAGTCTCACTATAC 4349
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 974 CCACAAACACACACACACACACACACACACACACACACACACACACACACAC 1033
XX
QY 4350 TTACGAGAGAGGACATGTGAACTTAATTAAGAGAGCTGACTACCTCAAAACATCCA 4409
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1034 CTACACACACACACACACACACACACACACACACACACACACACACACAC 1093
XX
QY 4410 AGACAAAGCTGGCAGATTTGAGAAATATACATTTGCTGGAAATTCAGACCTGTCTAC 4469
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1094 CAACAAACACACACACACACACACACACACACACACACACACACACACAC 1153
XX
QY 4470 CGATTATTAACACACACACATCTCC 4492
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1154 CAACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1176
XX
RESULT 55
ABT04778
ID ABT04778 standard; DNA; 5163 BP.
XX
XX AC ABT04778;
XX
XX 27-SEP-2002 (first entry)
XX
XX C parvum GP900 gene fragment SEQ ID NO: 4.
XX
XX Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
XX gene; ds.
XX
XX OS Cryptosporidium parvum.
XX
XX PN WO200194631-A1.
XX
XX PD 13-DEC-2001.
XX
XX PF 14-MAY-2001; 2001WO-US15624.
XX
XX PR 06-JUN-2000; 2000US-0588995.
XX
XX PA (REGC ) UNIV CALIFORNIA.
XX
XX PI Petersen C, Barnes DA, Nelson RG, Gut J;
XX
XX WPI: 2002-566447/60.
XX
XX Detecting Cryptosporidium in biological and environmental samples and
XX diagnosis of cryptosporidiosis involves, contacting the sample with
XX Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA
XX
XX Disclosure: Page 103-104; 157pp; English.
XX
XX The present invention relates to a method of detecting Cryptosporidium in
XX biological and environmental samples, and of diagnosing
XX cryptosporidiosis. This involves obtaining a sample and contacting it
XX with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
XX RNA, or its variant, mutant or fragment. The method is also useful for
XX detecting and identifying individual cryptosporidium isolates based on
XX the genetic characteristics, and for diagnosis of prior or concurrent
```


CC Cryptosporidium infection. The present sequence is a C. parvum coding
CC sequence used in the exemplification of the invention.
XX
XX Sequence 5163 BP: 1873 A; 1138 C; 875 G; 1277 T; 0 other:

Query Match 1.28; Score 60.6; DB 24; Length 5163;
Best Local Similarity 44.28; Pred. No. 0.00053;
Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

```
OY 3930 CAATGTCACGTCGAGCATTTCGACAGCGCGCTTTAATAAGAGAGAAATTTCAAGAA 3989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 614 CAACACACACACACACACACACACACACACACACACACACACACACACACACACAC 673
OY 3990 CGCTTTGAGAGCGACTGATTTCCTATTAAGACCTGACCTCCTCTACTACTGTTGCA 4049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 674 CAACACACACACACACACACACACACACACACACACACACACACACACACACAC 733
OY 4050 GGACCCATACAGTTCCTCAATTAAGAGAGAGGATTAATTTCTGTCTAGTTACACTA 4109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 734 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793
OY 4110 TATCAAGAGCTGCATCGACCGATATTCATTTGCTATATGATGCTGGGTATTACAA 4169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 794 CAACACACACACACACACACACACACACACACACACACACACACACACACACAC 853
OY 4170 GATCCTGAGCGAATAATACAGTCTGACCTTGACTCTATTAAAAAGTATATCAACGATTA 4229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 854 CCACACACACACACACACACACACACACACACACACACACACACACACACACAC 913
OY 4230 GCAGGCGAGAAATGAAAATATCTGCCCTTCTGTAATACATCGAAACCTCTGACAGAC 4289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 914 AGAAACCAACACACACACACACACACACACACACACACACACACACACACAC 973
OY 4290 AGTGAACGACAAATGACCTCTTCTGTAATTCACCTGGAGCGACAGTCTCAACTATAC 4349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 974 CCACACACACACACACACACACACACACACACACACACACACACACACACACAC 1033
OY 4350 TTACAGAGAGCAATGTGGAAGTTAAATCAAGAGCTGAATCTCAAAAACAAATCCA 4409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1034 CTACCACACACACACACACACACACACACACACACACACACACACACACAC 1093
OY 4410 AGACAGCTGGCAGATTTCAAGAAAATTAACAATTTGCTGGAATTTGACAGACTCTTAC 4469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1094 CAACACACACACACACACACACACACACACACACACACACACACACACACAC 1153
OY 4470 CGATTATACCAACAATCTCC 4492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1154 CAACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1176
```

RESULT 56
AAA61848
ID AAA61848 standard; DNA: 5318 BP.
XX
XX AAA61848:
AC
XX
XX 28-OCT-2000 (first entry)
DT
XX
XX DNA encoding a portion of Cryptosporidium parvum NINC isolate GP900.

XX
XX GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;
KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW merozoite; diarrhoea; protozoicide; ds.
XX
XX

OS Cryptosporidium parvum.

XX
XX Key Location/Qualifiers
FH 1669..7182
FT
FT CDS
FT
FT

FT /product= "Cryptosporidium parvum NINC isolate GP900"
FT /note= "No start codon given in the specification"
FT 3'UTR 5167..5318

FT /*tag= b

XX US6071518-A.

XX 06-JUN-2000.

XX 12-SEP-1997; 97US-0928361.

XX 13-SEP-1996; 96US-0026062.

XX 01-JUN-1993; 93US-0071880.

XX 29-MAY-1992; 92US-0891301.

XX 03-APR-1995; 95US-0415751.

XX 14-AUG-1996; 96US-0700651.

XX (REGC) UNIV CALIFORNIA.

XX Petersen C;

XX WPI; 2000-422065/36.

XX P-PSDB; AAB11727.

XX New GP900 protein fragments and fusion proteins of Cryptosporidium

XX parvum, useful for detecting the presence of the parasite, and

XX diagnosing or treating Cryptosporidium infections by competitive

XX inhibition of the function of GP900 -

XX Claim 16; Column 41-48; 59pp; English.

XX The invention relates to the GP900 glycoprotein of the protozoan

XX Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion

XX proteins comprising GP900 fragments. The invention also relates to the

XX administration of GP900 or fragments thereof to a host to elicit anti-

XX GP900 antibody production, and to a method of cryptosporidiosis treatment

XX or prophylaxis comprising administration of anti-GP900 antibodies to an

XX individual. Cryptosporidium parvum GP900 and GP900 fragments are able to

XX competitively inhibit sporozoite or merozoite attachment or invasion, and

XX are also useful for the generation of anti-GP900 antibodies. The

XX antibodies also inhibit sporozoite or merozoite attachment/invasion and

XX additionally inhibit the binding of GP900 ligands to GP900. GP900

XX proteins, fragments and antibodies may therefore be used to treat or

XX prevent cryptosporidiosis. Infection with Cryptosporidium is a common

XX cause of diarrhoea in humans and causes life-threatening diarrhoea in

XX immunocompromised persons. Cryptosporidiosis can be contracted from

XX contaminated municipal water supplies (e.g., public swimming pools). It

XX is also a cause of disease in animals, resulting in financial losses in

XX agriculture. GP900 fragments, fusion proteins and antibodies may also be

XX used for the diagnosis of cryptosporidium parvum infections, and for the

XX detection of the parasite in the environment. The present sequence

XX represents genomic DNA encoding a portion of the GP900 protein of the

SO Sequence 5318 BP: 1939 A; 1158 C; 890 G; 1331 T; 0 other:

Query Match 1.28; Score 60.6; DB 21; Length 5318;
Best Local Similarity 44.28; Pred. No. 0.00054;
Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

OY 3930 CAATGTCACGTCGAGCATTTCGACAGCGCGCTTTAATAAGAGAGAAATTTCAAGAA 3989

DB 614 CAAC 673

OY 3990 CGCTTTGAGAGCGACTGATTTCCTATTAAGACCTGACCTCCTCTACTACTGTTGCA 4049

DB 674 CAAC 733

OY 4050 GGACCCATACAGTTCCTCAATTAAGAGAGAGGATTAATTTCTGTCTAGTTACACTA 4109

DB 734 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793

OY 4110 TATCAAGAGCTGCATCGACCGATATTCATTTGCTATATGATGCTGGGTATTACAA 4169

DB 794 CAAC 853

[illegible]

Query Match	1.28;	Score 59.6;	DB 22;	Length 489;
Best Local Similarity	51.98;	Pred No. 0.00028;		
QY 1332	GAAATCTACACACATAGAGAAAGAGTTCATCAACGAAATCCAAAGAGATCAA	1391		
Db 2712	AAATTCAAAAAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2653		
QY 1392	AATTGAGACAGAGAAATTTAGAGTGACAAAGAAAGTTATGCAAGACCGCAAAAGTCT	1451		
Db 2652	AAAAAAAAACAAAAAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2593		
QY 1452	AAACGATATCAGTAAAGAGTATGAAAGC	1480		
Db 2592	AAACGAAAAAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2564		
RESULT 59				
AA500654				
ID	AA500654	standard; DNA; 489 BP.		
XX				
AC	AA500654;			
XX				
DT	07-SEP-2001	(first entry)		
XX				
DE	Plasmodium vivax merozoite surface protein DNA.			
XX				
KM	Merozoite surface protein; malaria; blood; serum; diagnosis; vaccine; ds			
KM	antigen-antibody composite; Enzyme Linked Immunosorbent Assay; ELISA;			
KM	MSP.			
XX				
OS	Plasmodium vivax.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	1..489		
FT		/*tag= a		
FT		/partial		
FT		/product= "P. vivax merozoite surface protein"		
FT		/note= "No start or stop codon"		
PN	W0200136587-A2.			
XX				
PD	25-MAY-2001.			
XX				
PF	15-NOV-2000; 2000MO-KR01302.			
XX				
PR	15-NOV-1999; 99KR-0050616.			
PR	25-APR-2000; 2000KR-0022041.			
PR	20-MAY-2000; 2000KR-0027305.			
XX				
PA	(HUMA-) HUMANBIO CO LTD.			
XX				
P1	Park H;			
XX				
DR	WPI: 2001-343809/36.			
DR	P-PSDB: AAU00668.			
PT	New gene encoding merozoite surface protein of Plasmodium vivax, useful			
PT	for producing protein for diagnosis of malaria and for vaccination -			
XX				
PS	Claim 1; Page 21-22; 24pp; English.			
XX				
CC	The sequence represents a DNA encoding a Plasmodium vivax merozoite			
CC	surface protein (MSP). The C-terminal region of the merozoite surface			
CC	protein has a strong antigenicity in malarial diseases. For diagnosis of			
CC	malaria, recombinant proteins with enhanced antigenicity, obtained by			
CC	addition of fusion proteins to surface protein C-terminal regions, can be			
CC	reacted with serum or blood of a Plasmodium infected patient.			
CC	Antigen-antibody composites will be formed, and these are detected by			
CC	Enzyme-Linked Immunosorbent Assay (ELISA). The recombinant antigens			
CC	provide a quick and reliable diagnosis of malaria, with good sensitivity			
CC	and selectivity.			
XX				
SO	Sequence 489 BP; 180 A; 134 C; 95 G; 80 T; 0 other;			

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 00:07:15 ; Search time 125 Seconds
(without alignments)
12119.866 Million cell updates/sec

Title: US-09-269-874A-2

Perfect score: 4940

Sequence: 1 cgcacgcgtatgaaatcatc.....ttcatctaatagatcatgtg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1265.2	25.6	5181	1	US-08-257-073-10 Sequence 10, Appl
2	331.6	6.7	1219	4	US-08-195-705-1 Sequence 1, Appl
3	74.2	1.5	7218	1	US-08-232-463-14 Sequence 14, Appl
4	60.6	1.2	5163	3	US-08-700-651-1 Sequence 1, Appl
5	60.6	1.2	5163	3	US-08-928-361B-4 Sequence 4, Appl
6	60.6	1.2	5318	3	US-08-700-651-2 Sequence 2, Appl
7	60.6	1.2	5318	3	US-08-928-361B-3 Sequence 3, Appl
8	58.8	1.2	5511	3	US-08-928-361B-2 Sequence 2, Appl
9	58.8	1.2	7334	3	US-08-928-361B-1 Sequence 1, Appl
10	50	1.0	5661	4	US-08-938-105-2 Sequence 2, Appl
11	49.6	1.0	1430	1	US-08-276-452A-25 Sequence 25, Appl
12	49.6	1.0	1430	2	US-08-798-744-25 Sequence 25, Appl
13	48.2	1.0	1690	1	US-08-276-452A-24 Sequence 24, Appl
14	48.2	1.0	1690	2	US-08-798-744-24 Sequence 24, Appl
15	46.4	0.9	198	5	PCT-US95-10668-3 Sequence 3, Appl
16	46.4	0.9	198	5	PCT-US95-10668-4 Sequence 4, Appl
17	46.4	0.9	954	4	US-08-098-327E-37 Sequence 37, Appl
18	46.4	0.9	954	4	US-08-462-625-37 Sequence 37, Appl
19	46.4	0.9	988	4	US-08-098-327E-34 Sequence 34, Appl
20	46.4	0.9	988	4	US-08-462-625-34 Sequence 34, Appl
21	45.8	0.9	340	1	US-08-182-175A-104 Sequence 104, App
22	45.8	0.9	340	5	PCT-US92-06412-104 Patent No. 5171843-8
23	45.6	0.9	697	6	5171843-10 Patent No. 5171843
24	45.6	0.9	1137	6	5171843-10 Patent No. 5171843
25	45.4	0.9	198	5	PCT-US95-10668-1 Sequence 1, Appl
26	45.4	0.9	198	5	PCT-US95-10668-2 Sequence 2, Appl
27	45	0.9	2949	4	US-09-412-554A-3 Sequence 3, Appl

28	44.8	0.9	950	4	US-08-098-327E-32 Sequence 32, Appl
29	44.8	0.9	950	4	US-08-462-625-32 Sequence 32, Appl
30	44.8	0.9	1482	4	US-08-098-327E-41 Sequence 41, Appl
31	44.8	0.9	1482	4	US-08-098-327E-45 Sequence 45, Appl
32	44.8	0.9	1482	4	US-08-462-625-41 Sequence 41, Appl
33	44.8	0.9	1482	4	US-08-462-625-45 Sequence 45, Appl
34	44.8	0.9	4765	5	PCT-US93-07261-10 Sequence 10, Appl
35	44.4	0.9	2277	1	US-08-676-967-2 Sequence 2, Appl
36	44.2	0.9	2277	1	US-08-676-967-2 Sequence 2, Appl
37	44.2	0.9	2277	2	US-08-676-967-2 Sequence 2, Appl
38	44.2	0.9	2277	2	US-08-098-487-2 Sequence 2, Appl
39	43.6	0.9	1956	4	US-08-559-896B-1 Sequence 1, Appl
40	43.6	0.9	1241	1	US-08-471-033-39 Sequence 39, Appl
41	43	0.9	1241	1	US-08-471-033-42 Sequence 42, Appl
42	43	0.9	1241	2	US-08-471-044-39 Sequence 39, Appl
43	43	0.9	1241	2	US-08-471-044-42 Sequence 42, Appl
44	43	0.9	1241	2	US-08-463-483A-39 Sequence 39, Appl
45	43	0.9	1241	2	US-08-463-483A-42 Sequence 42, Appl
46	43	0.9	1241	2	US-08-471-046A-39 Sequence 39, Appl
47	43	0.9	1241	2	US-08-471-046A-42 Sequence 42, Appl
48	43	0.9	1241	2	US-08-470-566B-39 Sequence 39, Appl
49	43	0.9	1241	2	US-08-470-566B-42 Sequence 42, Appl
50	43	0.9	1241	2	US-08-469-334-39 Sequence 39, Appl
51	43	0.9	1241	2	US-08-469-334-42 Sequence 42, Appl
52	43	0.9	1241	3	US-09-300-529-39 Sequence 39, Appl
53	43	0.9	1241	3	US-09-300-529-42 Sequence 42, Appl
54	43	0.9	1358	1	US-08-471-033-45 Sequence 45, Appl
55	43	0.9	1358	2	US-08-471-033-45 Sequence 45, Appl
56	43	0.9	1358	2	US-08-463-483A-45 Sequence 45, Appl
57	43	0.9	1358	2	US-08-471-046A-45 Sequence 45, Appl
58	43	0.9	1358	2	US-08-470-566B-45 Sequence 45, Appl
59	43	0.9	1358	2	US-08-469-334-45 Sequence 45, Appl
60	43	0.9	1358	3	US-09-300-529-45 Sequence 45, Appl
61	43	0.9	1389	1	US-08-471-033-27 Sequence 27, Appl
62	43	0.9	1389	2	US-08-471-044-27 Sequence 27, Appl
63	43	0.9	1389	2	US-08-463-483A-27 Sequence 27, Appl
64	43	0.9	1389	2	US-08-471-046A-27 Sequence 27, Appl
65	43	0.9	1389	2	US-08-470-566B-27 Sequence 27, Appl
66	43	0.9	1389	2	US-08-469-334-27 Sequence 27, Appl
67	43	0.9	1389	3	US-09-300-529-27 Sequence 27, Appl
68	43	0.9	4031	1	US-08-471-033-49 Sequence 49, Appl
69	43	0.9	4031	2	US-08-471-044-49 Sequence 49, Appl
70	43	0.9	4031	2	US-08-463-483A-49 Sequence 49, Appl
71	43	0.9	4031	2	US-08-471-046A-49 Sequence 49, Appl
72	43	0.9	4031	2	US-08-470-566B-49 Sequence 49, Appl
73	43	0.9	4031	2	US-08-469-334-49 Sequence 49, Appl
74	43	0.9	4031	3	US-09-300-529-49 Sequence 49, Appl
75	42.6	0.9	1493	4	US-08-098-327E-38 Sequence 38, Appl
76	42.6	0.9	1493	4	US-08-462-625-38 Sequence 38, Appl
77	42.6	0.9	1995	1	US-08-425-069-3 Sequence 3, Appl
78	42.6	0.9	1995	2	US-08-317-844B-3 Sequence 3, Appl
79	42.4	0.9	2447	2	US-09-014-969-14 Sequence 14, Appl
80	41.8	0.8	774	4	US-09-461-657-187 Sequence 187, App
81	41.8	0.8	819	4	US-09-461-657-185 Sequence 185, App
82	41.8	0.8	1669	4	US-09-461-657-184 Sequence 184, App
83	41.6	0.8	1236	2	US-08-741-134-5 Sequence 5, Appl
84	41.6	0.8	3312	4	US-08-923-992A-3 Sequence 3, Appl
85	41.6	0.8	3312	4	US-09-307-143-1 Sequence 1, Appl
86	41.4	0.8	1399	1	US-08-471-033-24 Sequence 24, Appl
87	41.4	0.8	1399	2	US-08-471-044-24 Sequence 24, Appl
88	41.4	0.8	1399	2	US-08-463-483A-24 Sequence 24, Appl
89	41.4	0.8	1399	2	US-08-471-046A-24 Sequence 24, Appl
90	41.4	0.8	1399	2	US-08-470-566B-24 Sequence 24, Appl
91	41.4	0.8	1399	3	US-08-469-334-24 Sequence 24, Appl
92	41.4	0.8	1399	3	US-09-300-529-24 Sequence 24, Appl
93	41.2	0.8	4411529	4	US-09-103-840B-1 Sequence 1, Appl
94	41	0.8	1360	6	US-08-961-083-37 Sequence 37, Appl
95	41	0.8	1866	6	5210183-1 Patent No. 5210183
96	41	0.8	12665	4	US-08-961-527-134 Patent No. 5210183
97	40.8	0.8	1131	6	5180810-3 Patent No. 5180810
98	40.8	0.8	1784	6	5180810-2 Patent No. 5180810
99	40.8	0.8	2058	2	US-08-749-391-1 Sequence 1, Appl
100	40.8	0.8	2058	3	US-09-390-200-1 Sequence 1, Appl

C 101	40.8	0.8	3407	1	US-08-253-155A-7
C 102	40.6	0.8	233	1	US-08-182-175A-86
C 103	40.6	0.8	233	5	PCT-US92-06412-86
C 104	40.4	0.8	188	1	US-08-115-457-21
C 105	40.4	0.8	188	1	US-08-466-670-91
C 106	40.4	0.8	188	2	US-08-291-011-1
C 107	40.4	0.8	188	4	US-09-266-065-1
C 108	40.4	0.8	405	2	US-08-299-074A-1
C 109	40.4	0.8	405	4	US-09-399-773-1
C 110	40.2	0.8	9636	1	US-08-323-170B-1
C 111	40.2	0.8	9636	4	US-08-954-441-1
C 112	40.0	0.8	4806	2	US-08-726-406A-185
C 113	40	0.8	3095	6	5231168-1
C 114	39.8	0.8	14442	3	US-08-781-891-208
C 115	39.6	0.8	1266	4	US-09-470-175-3
C 116	39.6	0.8	1796	4	US-09-470-175-1
C 117	39.6	0.8	4403765	4	US-09-103-840A-2
C 118	39.4	0.8	392	4	US-09-370-838B-18
C 119	39.4	0.8	2032	4	US-09-241-581B-5
C 120	39.4	0.8	2032	4	US-08-265-428-5
C 121	39.4	0.8	2032	5	PCT-US95-07721-5
C 122	39.2	0.8	720	3	US-09-172-063-17
C 123	39.2	0.8	720	4	US-09-316-919-18
C 124	39.2	0.8	768	3	US-09-172-063-33
C 125	39.2	0.8	3489	2	US-08-728-333A-1
C 126	39.2	0.8	3489	4	US-09-298-568-1
C 127	39.2	0.8	32207	2	US-08-770-379-20
C 128	39.2	0.8	32207	4	US-08-757-669A-20
C 129	39.2	0.8	32207	4	US-09-230-371A-20
C 130	39.2	0.8	535	4	US-09-056-556-171
C 131	39	0.8	535	4	US-09-072-556-165
C 132	39	0.8	696	4	US-09-461-687-193
C 133	39	0.8	699	4	US-09-461-687-191
C 134	39	0.8	717	4	US-09-461-697-189
C 135	39	0.8	12286	4	US-09-322-478-17
C 136	39	0.8	152331	3	US-09-128-155-16
C 137	38.8	0.8	1695	4	US-09-308-022-5
C 138	38.8	0.8	3337	1	US-08-072-610-1
C 139	38.8	0.8	3337	2	US-08-719-822B-1
C 140	38.8	0.8	4337	4	US-09-092-458-1
C 141	38.6	0.8	440	4	US-09-397-787-285
C 142	38.6	0.8	1621	2	US-08-933-750C-96
C 143	38.6	0.8	1621	3	US-09-234-613-96
C 144	38.6	0.8	2089	1	US-08-441-139-13
C 145	38.6	0.8	19124	2	US-08-448-826B-13
C 146	38.4	0.8	240	1	US-08-628-417-6
C 147	38.4	0.8	2395	4	US-09-319-989-5
C 148	38.4	0.8	2395	4	US-09-319-989-7
C 149	38.4	0.8	3081	4	US-09-319-989-9
C 150	38.4	0.8	3360	4	US-09-319-989-8
C 151	38.4	0.8	3384	4	US-08-923-992A-5
C 152	38.4	0.8	4200	1	US-08-242-932-1
C 153	38.4	0.8	4200	1	US-08-847-065-24
C 154	38.4	0.8	4200	4	US-08-714-481-1
C 155	38.4	0.8	4200	5	US-08-923-992A-1
C 156	38.2	0.8	1596	4	PCT-US95-06111-1
C 157	38.2	0.8	2462	4	US-09-134-001C-2083
C 158	38.2	0.8	3023	4	US-08-847-065-24
C 159	38.2	0.8	5394	3	US-09-308-022-4
C 160	38	0.8	289	4	US-08-688-376-1
C 161	38	0.8	289	4	US-09-007-005-17
C 162	38	0.8	289	4	US-09-244-796-17
C 163	38	0.8	321	4	US-09-0

Sequence 7,	Appl
Sequence 86,	Appl
Sequence 21,	Appl
Sequence 1,	Appl
Sequence 1,	Appl
Sequence 1,	Appl
Sequence 1,	Appl
Sequence 1,	Appl
Sequence 20,	Appl
Sequence 20,	Appl
Sequence 171,	Appl
Sequence 166,	App
Sequence 193,	App
Sequence 189,	App
Sequence 189,	App
Sequence 16,	Appl
Sequence 5,	Appl
Sequence 1,	Appl
Sequence 1,	Appl
Sequence 285,	Appl
Sequence 96,	Appl
Sequence 13,	Appl
Sequence 6,	Appl
Sequence 5,	Appl
Sequence 1,	Appl
Sequence 1,	Appl
Sequence 1,	Appl
Sequence 2083,	Ap
Sequence 4,	Appl
Sequence 1,	Appl
Sequence 17,	Appl
Sequence 17,	Appl
Sequence 159,	App
Sequence 22,	Appl
Sequence 3,	Appl
Sequence 5,	Appl
Sequence 9,	Appl
Sequence 10,	Appl
Sequence 6,	Appl
Sequence 6,	Appl
Sequence 3,	Appl

174	38	0.8	3153	4	US-09-695-782-3	Sequence 3, Appl
175	38	0.8	3694	4	US-09-080-625-5	Sequence 5, Appl
176	38	0.8	3694	4	US-09-695-782-5	Sequence 5, Appl
177	38	0.8	5552	3	US-08-155-888-1	Sequence 1, Appl
c 178	38	0.8	9646	3	US-08-811-566-1	Sequence 1, Appl
c 179	38	0.8	9646	4	US-09-034-756-1	Sequence 1, Appl
180	37.8	0.8	207	1	US-08-700-575-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1

US-08-257-073-10

Sequence 10 Application US/08257073

Patent No. 5766597

GENERAL INFORMATION:

APPLICANT: Paolletti, Enzo

APPLICANT: de Taisne, Charles

APPLICANT: Tine, John A.

TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtiss, Morris & Safford, P.C.

STREET: 530 Fifth Avenue, 25th Floor

CITY: New York

STATE: New York

COUNTRY: UNITED STATES OF AMERICA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/257,073

FILING DATE: 09-JUN-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/075,783

FILING DATE: 11-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/852,305

FILING DATE: 18-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/672,183

FILING DATE: 20-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454310-2570

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

TELEX: 425066 CURTMS

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 5181 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-257-073-10

Query Match

Best Local Similarity 25.6%; Score 1265.2; DB 1; Length 5181;

Matches 2912; Conservative 0; Mismatches 1993; Indels 291; Gaps 16

QY 10 ATGAATATCATTTCTCTGCTGTTCAATTCGTGTTTTATATCATCAATCTAGTCGTG 69

|||||

Db 1 ATGAGATCATATTCCTTCTATGTCATTTCTTTCTTATATAAATACACAATGTGA 60

|||||

QY 70 ACCCAGATCTCTATCAGAGCTGTGTAGAAACGTGAGCGCTTGGAATATCCCGCTT 129

|||||

Dh 61 ACACATGAAAGTTATCAAGAACTGTCAAAAACTAGAGCTTTAGACATGCGAGTATG 120
Oy 130 ACCGATACAGCGCTGTCCAGAGGAGAGATGCTCAATGAA----- 174
Dh 121 ACAGGTTATGGTTTATTTCATAGGAAAAAATGATCTTAAGTAAGAAATAATTACTACA 180
Oy 175 -----GGACAGAGTGGCAGCGCGCTTACAAACAGCAGCACCGCT 213
Dh 181 AAAGTGCAAGTCTCAAAAGTGTACAAAGTGTACAAAGTGTACAAAGTGTACAAAGTGT 240
Oy 214 TCTAAAGGCTGTGTGTAGCGGTGCGTCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 273
Dh 241 ACAAGGAGTACAAAGT 300
Oy 274 GGCTCCGGGAGCGGT 333
Dh 301 GGTACAGAGT 360
Oy 334 CGAAGAACCAATCATCTGTACA----- 355
Dh 361 AGTCATCATCTGTGTCAAAACCTTACCTGTGTCAAAATCTGTGTGTGTGTGTGTGTGTGT 420
Oy 356 -----ACTTAGCGATTCGACGCCAAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 408
Dh 421 CCAGCTGATGCAAGCGATTCAGATGCTTAATCTTACGCTGATTTAAACACAGAGTACGA 480
Oy 409 AACTATCTCTGACATATCAAGAGGTGAAGTACCCAGATGTGTGTGTGTGTGTGTGTGTGTGT 468
Dh 481 AATTACTTGTACATTTAAAGAACTCAATATCCGAACTCTGTGTGTGTGTGTGTGTGTGTGT 540
Oy 469 ATGTGACACGTGTGTATTAACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 528
Dh 541 ATGTAACTTGT 600
Oy 529 ATCAATGAACCTCTGT 588
Dh 601 AATTAATGAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 660
Oy 589 GAGCTTGGCCAAATGACTATGTCAAAATTCATTCATTAATTAATTAATTAATTAATTAAT 648
Dh 661 GATGTATGT 720
Oy 649 TTGACGATTTGAGAAAGT 708
Dh 721 TTAGACGTACTTAAAAAAGT 780
Oy 709 AATGTGGGAAAGT 768
Dh 781 AATGTAGGAAAAATGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
Oy 769 GAGCTGATGAGAAATTCAAAAAGACATGACAAAAATTAAGAAATTAAGAAATTAAGAAAT 828
Dh 841 GATTTATTTAGGAAGATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 900
Oy 829 GAAAGAAAGAAAGT 888
Dh 901 GAAAAAAGAAAGT 960
Oy 889 GAGGCCATTAACCTGATGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 948
Dh 961 GAGGACATTAATTTAATAAGCTTTTAAAGAAACGATTTAAGAACTTTAAGAAATTAAGAA 1020
Oy 949 AATATCAAAAGATCTGT 1008
Dh 1021 AACATTAAGATTTACTGT 1080
Oy 1009 GGGAAACCCCTTAACAGCTGT 1068
Dh 1081 GGAATAATACCAAAATACTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
Oy 1069 ATCAAAAGAGATGCGCAAAAGCTTAAGTTCAACATGATGTCTCTTACTGTATCCCTT 1128
Dh 1141 ATMAAAGAAATTTGCCAAAACTTTAAATTTTACATTTGATGATTTTACTGTATCCACTT 1200

Oy 1129 GAGCTGAGTACTACTGTGAGAGAAAGTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 1187
Dh 1201 GAATTAAGATTAATTTAAGAAAAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 1260
Oy 1188 -----GACAAGAAATCAACCGAAGCTTAATGAATATCCAAATGTGTGTGTGTGTGTGT 1239
Dh 1261 GATCTCAGAAATTCGTTCAAATACCAAAAGTCTCTATCCAAATGTGTGTGTGTGTGTGTGT 1320
Oy 1240 GTGTCTTATTAACGATATCAACA-----CGCTCTCAACGAGCTCAATAGCTTGTGTGT 1290
Dh 1321 TTACCACTGACTGATATTCATATTCATATGCTGAGATTAAGAAATTAAGAAATTAAGAAAT 1380
Oy 1291 GACTTGATTAACCCCTGT 1350
Dh 1381 GATTTAATGAATTCGT 1440
Oy 1351 GAGAGAAAGAAATTAATCAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 1410
Dh 1441 GAAAGAAAAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
Oy 1411 GAGAGTGACAAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 1470
Dh 1501 AATCAACAAAAAGAAACAAAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAA 1560
Oy 1471 TATGAAAGCTGT 1530
Dh 1561 TATGAAAGATTTCTGT 1620
Oy 1531 AACTTCGAGAAATGATGT 1590
Dh 1621 GT 1680
Oy 1591 AATACCTTGT 1650
Dh 1681 AATTAATTTTCACTGT 1740
Oy 1651 AAGTATATGAGAGATTTCTGT 1710
Dh 1741 TCATATCTTGAAGATTTATTTCTTAAAGAAAGAAATTTCTGAAAAAGATTTTAAATCATTT 1800
Oy 1711 AAGAAATCTAATGAATTAAGATGAAAAAGAGATGAGACCTTGTGTGTGTGTGTGTGTGTGT 1770
Dh 1801 TATACTTTGAAAAAGT 1860
Oy 1771 GATTAAGAAACACTGT 1830
Dh 1861 AGTGAAGAAACAAATTTCTAGAAAAAATTTTAAAGAACTTAACATTCAGCAATG--CT 1917
Oy 1831 ATCTGTGAGGTCTCGATTTGTAAAGTCCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1890
Dh 1917 TCCTTGAAGATTAATGATTTGTAAATTTACAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1977
Oy 1891 AATGATGAATCAAGAAAGCTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1950
Dh 1978 ATAGAAGAACTTAAGAAAGTAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTT 2037
Oy 1951 CATGTCCGAATTAATTAAGAAAGTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 2010
Dh 2038 CATGTACCAAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTT 2097
Oy 2011 AAGAAAGATATGACAAATCTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2070
Dh 2098 AAAAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2155
Oy 2071 GAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2130
Dh 2156 AAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2215
Oy 2131 GAGATTAACCGGACAGGCTTACCAACCAAGCCGCAACAGGCGGTTCAGCTGTGCAAGGC 2190
Dh 2216 AAGAT--GGGGGTCACTCCACACACATTTATCCAAATGAGGAAAGAAAGAAATTAACAGAA 2274

Oy	2191	GATAGCTGCAGCCTGACAGCAAGAGCAAGAGCAGGCAACGACCTCCAGCTGCAGTGCAGTGC	2250
Db	2275	GAAACAGAGAAGAACAGAGAAACAGTAGAGCACACACACAGCGGTAAACATTAACATTACCA	2334
Oy	2251	GTTCACAGAGCCTAAAGCTCAAGTGCCTACACACACAGCTCCTGTGAAATACAGACCGAG	2310
Db	2335	CCAAACAGAAAGTAAAGTTGTTGAAGAAATTCATATAGAACATAGAGTATGACATTTCACA	2394
Oy	2311	AATGTACAGCAATCGACTACCTTGAGAAAGCTCTATGAGTCTCTGAATACATCCTACATC	2370
Db	2395	GCCCTTGACAAAACAGCTTTATCTTAAGAAATTAGATGAATTTTAACTAAATCATATATA	2454
Oy	2371	TGGCACAATATATCCCTGCCTCTCCACACACTATGACAAAGAAATCTTAAACAGTAC	2430
Db	2455	TGTCATAAATATATTTAGTATATCAAACTCTAGATGTGACCAAAAATTTATGAGAGTAT	2514
Oy	2431	AAGATTAACCAAGAGAGAGAGAGTAAACTGCTCTCTGTGATCCACTGAGACTGCTGTC	2490
Db	2515	AATCTTACTCCAGAGAGAGAAATGAAATTAATATCATGTGATCCATTAGATTTATATT	2574
Oy	2491	AATATCAGAAACATTTCCCGTTATGTATCTATGTTCGATAGCCTCAACAAATTCCTC	2550
Db	2575	AATATCAAAATATACATACCTGCTGATGTATTCATTTATGATAGTATGACAAATGATTTA	2634
Oy	2551	TCTCAACTGTTCAATGAGATATATGAGAAAGAGATGCTGCACACTGTATTAACCTCAA	2610
Db	2635	CAACATCTCTTTTGTGAATTAATTAACAAAGAAATGATTTATTTACATTAACCTAATA	2694
Oy	2611	GACACACACAAGATTATAGACCTCTTCGAGAGAGAAAGCTACAGAGAGCTC-----	2657
Db	2695	GAGAGAAATCACATCAAAAATTTATAGAGAGACAAAACAAATATACGTGAAACATCATCT	2754
Oy	2658	-----CACCTCTGTTAAACTCTCTCTTC	2681
Db	2755	ACATCCAGTCCCTGGAAATACAACTGTAATATCTGCTCATCCGCAACTCACAGTAAATTC	2814
Oy	2682	CAGCTCATGCAACACACTGTCTCTCACACTCAGACAA-----	2720
Db	2815	CAAAACACACATCAAAATGATCTCTCTACCAATACCCAAATATGTTGACTGTATCATCT	2874
Oy	2721	-----GCCCGAAGTGAAGCGCTTAACGAGACACCTCTCACTCGACCAACCTTAATACCTCA	2775
Db	2875	GGTCTCTGTAGTTGTAAGAAAGTCAATGATCCTTAACAGTATGTGTATATAGTACGAT	2934
Oy	2776	CTGAACATGTTTGAGAACATCTGTCTCTCGCAGAGATTAAGACAT---CTACCAAGAA	2832
Db	2935	TTGGAAGGTATTTGTACTCTCTTAATCTTTGGAATTAACCTTAAGTACTTAATCATTTA	2994
Oy	2833	CTTTATGGAACGAATGTGCTCGAAGCTCTCTGACAAATATCTGAAAGACAGCAGCA	2892
Db	2995	ACCAATTTCTACACAGATGAGAAAAATTTTATGAGAAATATTTTAAAAATATATATACC	3054
Oy	2893	TTCTATATACGAGACTCTCACTAACTCTTGTAATCTTAAGACCGAGATATCAACTCTCT	2952
Db	3055	TATTTTATATGATATATCAACAACTTCGTAAATCTTAATTTCAAAAGTATATTAACAGTTG	3114
Oy	2953	AAGCATGATCTAAACGTAAAGAGCTGGAAGAGACATCAATTAAGCTGAAGAAGACACTG	3012
Db	3115	ACC---GAAACACAAAAAATGCAATTAATATGATGAATTAATAAATTAAGATACTTTA	3171
Oy	3013	CAACTGAGCTTCGACCTGTACACAAAGTACAAACTGAAGAGACTCTTGACAGAG	3072
Db	3172	CAGTATCATTTGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3231
Oy	3073	AAGAAGACAGTCCGCAAGTATTAAGATGACAGATCAACATCTGCTGCTCAAGAGAGAG	3132
Db	3232	AAAAAGCACTTGCCCAAGACAAAATGCAATTAATAAATACTTACTTTATTAAGAAGACAA	3291
Oy	3133	CTTGAAGCAAACTCACTACTGAACAAATCCGAAACAGTACTGACAGAACTCTCACTG	3192
Db	3292	TTAGATATCAAAATGTAATCTACCTTAATTAACCCACATTAATGTATTTACAAAACCTTTCTGTT	3351
Oy	3193	TTCTTTCACAAAGAGAGAGAGCCGAGATGCCGAGACAGAGAACCTCTGTGAGAACACC	3252

Db	3352	TTCTTTAAACAAAAAAGAGAGCTGAATATAGAGAAACCTGAAACACATTTAGAAAGACA	3411
Oy	3253	AAGATCTTCTCAAAACACTACAAAGCCCTGTCAGATATTATAATGCGAGCTCTTCTCT	3312
Db	3412	AAAATATATATTAAGAACCTTTTAAAGAGACTGTGTAATATTAATAAAGTGAAATCATCTCA	3471
Oy	3313	CTGAAGACTCTCCCGAGGAGACATCCAGACCAGAGATATACCGCCAGCCTGAGAAC	3372
Db	3472	TTAAAAACCTTAAAGTGAAGATATCAATTTCAACAGAAAGATATATTAGCCATTTTAGAAAA	3531
Oy	3373	TTCAAGTCTGTCTAAGCTCGAAGGCAAGCTGAAGGACAACTGAGACCTGAGAAAG	3432
Db	3532	TTTAGAGATATTAAGTAAATATAGATGAAAACTCAATGATTAATTTACATTTTAGAAAGAA	3591
Oy	3433	AAGCTCAGCTACCTCTTACCGGACATCATCTGATTCGCCAGCTCAAGGAAGTCTT	3492
Db	3592	AAATATCTTTCTTATCACTGATTAACATCATTTTAATTAAGCAATTTAAAGAAATATA	3651
Oy	3493	AAGAAACAAGACTACACCGGCAATATACCCAGCGAAGATTAATACAGCTGATTAACGA	3552
Db	3652	AAAATTAATAATTTATACAGTATTTCTCCAGTGAAGAAATATATACAAAGTTACAGAACT	3711
Oy	3553	CTGAATCTTACAAAGATTCCTGCTGAAGGAAACAGATGTGCG-----	3596
Db	3712	TTAAAACTTACGAAATTTTCTCCAGAGCAAAAGTTACAAAGTTGATTAATCTCACT	3771
Oy	3597	-----CACTGTGGTGTGTAATGTGCTCCGACACATGAGAGACTGTCA	3641
Db	3772	CAACCAGATGTAATCCATCTCCATTTATCTGTAAAGGTAAAGTGTACAGATCCACA	3831
Oy	3642	ACCTAAGAA-----GCTGTGATCTACTCATGTGCGAGCCGAGTCCATTAACAATTAC	3693
Db	3832	AAAGAGAAACACAAATACCACTTCAGGCTCTTTATTACGAATTTACAAACAGTACTA	3891
Oy	3694	ACATCTAGAAGCTGACGAGATGAGGTGACATGATCATATGTGCTATCTGTGGGAG	3753
Db	3892	CAATTACAAATTTATGCGAGAGAGATGATCTTATGTTGTATTTACCATTTTGTGAGAA	3951
Oy	3754	AGCGAGGAGCATACGATGACCTGGCCAGAGTGTCACCGTAGGCTGTCACTCTTCC	3813
Db	3952	TCCGAAGTATATGCGAATTTTATGATCAAGTATGTAACGTGAGGAAGC---AATATCTGTC	4008
Oy	3814	GTGATGTATACATCTGTGCCAAATTCGAAACGAATACGAAGTCTCTATCTGAAACT	3873
Db	4009	ACAATGATATATATCTCTAGGATTTGAAATGAAATGATGATATATATTAATAACT	4068
Oy	3874	CTGGAGGCGCTGTAGTGTCTCAAGAAACAGCTGGAATATAGCTGATCACTTCAAT	3933
Db	4069	TTAGCTGAGATATATACAACTTAAAAAACAATTTGAAAAAACATTTTACATTTAT	4128
Oy	3934	GTCAACGTGAAGCATTTGTGAACAGCCGCTTTAATACAGAGAAATTTCAAGACCT	3993
Db	4129	TTAAATTTGAACGATATCTTAAATTCACGCTTTTAAACGAAAAATTTTCTAGATGA	4188
Oy	3994	TTGGAGGCGCATGATTTCCCTTAATTAAGACCTGACCTCTATACCTAGCTGTCAAGAC	4053
Db	4189	TTAGATCTGATTTAATGCAATTTAAACATATATCTCTCAATGAATACATTTTGAAGAT	4248
Oy	4054	CCATACAGTTCCTCAATTAAGAGAAAGAGGATAAATTTCTGTAGTTACAATATATC	4113
Db	4249	TCATTTAAATATATGATTTAGAAACAAAAAACACACTTTTAAAGATACAAATATATA	4308
Oy	4114	AAGAGCTCATCCGACACCGATATATCTTCGTAAATGATGTCTGGGGATATACAAAGTC	4173
Db	4309	AAAGATCTAGTAAATGATTTAAATTTGCAACAGGAAGGTATTAATTAATGAAGAAAG	4368
Oy	4174	CTGAGCGAAAAATACAGTCTGACCTTGACTATTAATAAATATATATCAAC-----	4224
Db	4369	GTTTATAGCGAAATATATAGATGATTTAGATCAATTTAAAAAAGTTATCAAGAAAGAAAG	4428
Oy	4225	-----GATATACAA	4233

Db 4429 GAGAGATTCCATCATCAACCAACACACCTCCGTCACGACAAAAACAGACGACAA 4488
QY 4234 GCGGAGATGAAATATTCCTGCCCTCCGATATACATCGAAACCTGTACAGACAG 4293
Db 4489 AAGAAGAAAGTAAAGTCTCCATTTTAAACAAACATTGAGACCTTATACATACTTA 4548
QY 4294 AACGAAATATGACCTCTTCCTAATTCACCTGAGAGCCAGGCTCCTCACTATCTTAC 4353
Db 4549 GTTAAATTAATGACGATTACTTAATTAACGCAAAATTAACGATTGTATGT 4608
QY 4354 GAGAAAGCAATGTGGAAGTAAATCAAGAGCTGAACCTCAAAACAAATCCAGAC 4413
Db 4609 GAAAGATGAGACACATGTTAAATTAATACTAATAGATTAAACCAATGATGAC 4668
QY 4414 AAGCTGGCAGATTTCAGAAAAATTAACAATTCGTGCGAATTCGACAGCTGTACCGAT 4473
Db 4669 AAAATGATCTTTTAAAAACCATTAACGCTCGAAGCAATTAATAATGATTAATGAT 4728
QY 4474 TATTAACCAACAAATCTCTGACCAAGTTCTGTCACAGCTGCGATGGTGTCCAAAACTC 4533
Db 4729 GATAGAAAAAAGATATGCTTGCAATTAATTAAGATGATGATGATGATGATGATGAT 4785
QY 4534 GCGAAACAGCTCTGACGATGCTGACGCGCAACCTGACAGGATGCTGAACATCTCC 4593
Db 4786 CCTAATACATATATCAAAATTAATGAAGAAAAATTCAGATATGTTAAACATTTCA 4845
QY 4594 CAGCAACATGCGTGAAGAAAGAGTCCGCAAGTATGAGGCTGTTGAGGATCTGAC 4653
Db 4846 CACACCAATGCGTAAAAAACAAATGTCGAAATTCGATGTTTCAGACATTTAGAT 4905
QY 4654 GAGCGCAGAGTGTGCAAGTCTCTGCACTCAAAACAGAGAGATGAGTGCCTGAG 4713
Db 4906 GAAAGAGAAATGTAATGTTTAAATTAATTAACAAAGAGTGTATTAATGTTGAA 4965
QY 4714 AACCAAACTCACTGCAATGAAACAAATGCGGCTGTGACCGCGATCTAATGACAC 4773
Db 4966 AATCCAAATCTTCTGTAACCAAAATATGATGATGATGATGATGATGATGATGATGAT 5025
QY 4774 GAGGAAGACAGCGCTGTAAGGAAAGAAATACATGATGATGATGATGATGATGATGAT 4833
Db 5026 GAGGAAGATTCAGGTAGCAACGAAAGAAATGATGATGATGATGATGATGATGATGAT 5085
QY 4834 TATTCATCTTCGAGCGGATTTTTCCTCAGCTTAATTTCTGGGATCTCTCTCTCTCT 4893
Db 5086 TATTCATCTTCGAGGTATTTTTCGAGCTCTCACTTTTGAATATATCTCTCTCTCT 5145
QY 4894 CTGATCTCATCTGATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 4929
Db 5146 TTAATATCTCATGTTAAT 5181

RESULT 2

US-08-195-705-1
; Sequence 1, Application US/08195705
; Patent No. 6420523
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra
; APPLICANT: Hui, George
; APPLICANT: Bair, Phillip
; APPLICANT: Gibson, Helen
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDUM
; TITLE OF INVENTION: FAICIPARUM VACCINE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: N.Y.
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/195,705
APPLICATION NUMBER: 6420523
FILING DATE: 14-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jacobs, Seth H
REGISTRATION NUMBER: 32140
REFERENCE/DOCKET NUMBER: 11880A3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-757-2200
TELEFAX: 212-586-1461
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Plasmidium faiciparum
STRAIN: faiciparum uganda palo alto
IMMEDIATE SOURCE:
CLONE: f195LFP42ACNPV
FEATURE:
NAME/KEY: exon
LOCATION: 13...1194
US-08-195-705-1

Query Match 6.7% Score 331.6; DB 4; Length 1219;
Best Local Similarity 57.9%; Pred. No. 7.3e-78;
Matches 669; Conservative 0; Mismatches 424; Indels 63; Gaps 2;
QY 3765 CTACATGACCTCGGCGGAGGTGTCACCGGTGAGGCTGTCACTCTCCGTGATTGATTA 3824
Db 39 CTGGGCTGTGCTGTGTCACGACGACACTGTGACCGGCGGATGTGTACAAATGATTA 98
QY 3825 CATTTGTCCAAATGAGAAATGCAAGTGTCTATCTGAAACCTGTGGCAGGCT 3884
Db 99 TATCTCTCAGGATTTGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 158
QY 3885 CTATAGTCTCTCAAGAAACACTGGAATAGCTGATGATGATGATGATGATGATGATGAT 3944
Db 159 ATATAGAGCTTAAAAAACAATGAAAAAATTTTACATTTAATTAATTTGAA 218
QY 3945 GACATTTGACACCGCTTAAATAGAGAAATTTCAAGACGTTGTGAGAGCGA 4004
Db 219 CGATATCTTAATTTCACTGCTTAAAGAAACGAAATTTTCTTAGATGATGATGATGAT 278
QY 4005 CTGATCTCTTAAAGAGCTGACCTCTTAAATAGAGAAATTTCAAGACGTTGTGAGAGCGA 4064
Db 279 TTTAATGCAATTTAAACATATATCTCAATGATGATGATGATGATGATGATGATGAT 338
QY 4065 CCTCATTAAGAGAGAGGATTAATTTCTGTAGTATGACATCTATCAAGAGACTCCAT 4124
Db 339 ATTGAATTCGAAACAAAAACACACTTTTAAAGATTAACAATATTAATAAAGATTCAGT 398
QY 4125 CGACACCGATATCAATTTGCGTAATGATGCTGGGATTTCAAGATCTCTGAGGAA 4184
Db 399 AGAAATGATTAATTAATTTGACAGAGAGTAAAGTTATGAAAAAGTTTATGCGAA 458
QY 4185 ATCAAGTCTGACCTTGAATTAATAAAGTATCAACGATTAAG-----CAAGCGAATGA 4244
Db 459 ATATAGAGTATTAATGATCAATTAATAAAGTATTAACAAGAGAGAGAGAGAGTCCC 518
QY 4231 -----CAAGCGAATGA 4244
Db 519 ATCATCACCAACACACCTCTGACACGCAAAACAGAGAGAGAGAGAGAGAGAGAG 578
QY 4245 AAAATATCTGCCCTTCTGATTAATGATGAAACCTGTATGAGAGAGAGAGAGAGAGAT 4304

DB 579 TAAGTTCCTTCATTTTAAACAACATTTAGACCTTATACAAATAGTATTATAAAT 638
QY 4305 CGACCTCTTCGTATTTACCTGAGGCCAAGTCTCACTATATCTTACGAGAAGACAA 4364
DB 639 TGACGATTACTTAAATTAACCTTAAGGCAAGATTAACGATTGTAATCTTAAGATGA 698
QY 4365 TGATGAAGTAAATCAAGAGGCTGACCTCACTCAAAACATCCAGACAGAGTGGCAGA 4424
DB 699 AGCAGATGTTAAATTAATCACTAATCTGATTTTAAAGCAATTTGATCAACAATAGATCT 758
QY 4425 TTTCAAGAAAAATPACAAATTTCTGCGAATTTGACAGACCTGTCAACGATTATACCAAA 4484
DB 759 TTTTAAAAAACATACAGACTTTCGAGCAATTTAAAAATTTGATTAATGATATACGAAAA 818
QY 4485 CATCTCTCCACCAAGTTTGTGTCACATGTCGATGTTTGGAAAACTCCCAAAACAGT 4544
DB 819 AGATATGCTTGGCAAAATTAAGTACAGAGATTAG--TTCAAAATTTTCTTAATCAAT 875
QY 4545 GCTGAGCAATCTGCTCGACGCAACCTGACGGCATGCTGAACATCTCCAGACCAATG 4604
DB 876 AATATCAAAATTAATGAAGAAATTCGAAGATATGTTAAACATTTTACACACCAATG 935
QY 4605 CGTGAAGAAACAGTCCCGCAGAAATAGCGCTGTTTCAGCATCTGACAGCGCGAAGA 4664
DB 936 CGTAAAAAAACATGTCGACAAATTTCTGATGTTTCAGACATTTAGATGAAGAGAGA 995
QY 4665 GTCAACAGTCTCTCGTAACAAACAAGAGAGATAGTGGCTGGAGAACCAACCC 4724
DB 996 ATGTAATGTTTAAATTAATTAACAACAAGAGATGATGTTTAAATTTCAAAATTC 1055
QY 4725 TACCTGCATGAAGAAACATGGCGGTGTGACGCCGATGCTAAATGACGAGAGACAG 4784
DB 1056 TACTTGAAGCAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1115
QY 4785 CGGCTTAACGGAAGAAATTAACATGCGAGTACTAAGCCGACCTGATCCAGCTT 4844
DB 1116 AGTAGCAAGCAAGAAATTAACATGATGATGATGATGATGATGATGATGATGATGATG 1175
QY 4845 CGACGGATTTTTCG 4860
DB 1176 CGATGATTTTTCG 1191

RESULT 3
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 999149
INFORMATION FOR SEQ. ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pl-Fls
US-08-232-463-14

Query Match 1.5%; Score 74.2; DB 1; Length 7218;
Best Local Similarity 7.88; Pred. No. 3,2e-09;
Matches 34; Conservative 234; Mismatches 167; Indels 0; Gaps 0;

QY 1693 AAGAACTAAAGTATTAACAAGATCTCATAGTAAGTCAAGAAACGAGATCGAGCGCTT 1752
DB 1465 AAGTAACTAAAGTATTAACAAGATCTCATAGTAAGTCAAGAAACGAGATCGAGCGCTT 1406
QY 1753 GTTGAGAACATTAAGAAAGTGAAGAACAGTGTGTTGAGAAAGATTACAAAAGACAA 1812
DB 1405 RRR 1346
QY 1813 AATAAACAGTAAAGAGATCTGAGAGTCTCGATATTGTTAAAGTCAAGTCAAGAG 1872
DB 1345 RRR 1286
QY 1873 GTGCTCTCATGAACAAGATTGATGAACCTCAAGAGCTCACTCATCTGGAAGAGCTG 1932
DB 1285 RRR 1226
QY 1933 GAGTTAAACATATATATATATGTCGCAATAGTTATAGCAGAGGAATTAAGCAAGCA 1992
DB 1225 RRR 1166
QY 1993 TACTACCTCATCTACTCAAGAAAGAGATAGACAACTGAAGTTCATGCCCAAAGTC 2052
DB 1165 RRR 1106
QY 2053 GAGAGCTGATCAAGCAAGAGAGAGACATTAATACTGAAGCAGTCAGATACTCC 2112
DB 1105 RRR 1046
QY 2113 GAGCCTTCACAGAA 2127
DB 1045 CAGCAAGCTCGGAA 1031

RESULT 4
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751

```

: EARLIER FILING DATE: 1995-04-03
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 5163
: TYPE: DNA
: ORGANISM: Cryptosporidium parvum
US-08-700-651-1

```

Query Match	1.2%	Score 60.6;	DB 3;	Length 5163;
Best Local Similarity	44.2%;	Pred. No. 1e-05;		
Matches 249; Conservative	0;	Mismatches 314;	Indels 0;	Gaps 0;

[illegible]

RESULT 5
US-08-928-361B-4
; Sequence 4, Application US/08928361B
; Patent No. 6071518

1 APPLICANT : Petersen, Carolyn
 2 TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
 3 TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
 4 TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
 5 TITLE OF INVENTION: SPECIES INFECTIONS
 6 NUMBER OF SEQUENCES: 30
 7 CORRESPONDENCE ADDRESS:
 8

```

:
:
: ADDRESSSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
:
: COMPUTER READABLE FORM:
:

```

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: PatentIn Release #1.0, Version #1.30
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/928.361B
7 FILING DATE: 12-SEP-1997
8 CLASSIFICATION:
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 60/026,062
11 FILING DATE: 13-SEP-1996
12 ATTORNEY/AGENT INFORMATION:
13 NAME: VERNY, HANA
14 REGISTRATION NUMBER: 30,518
15 REFERENCE/DOCKET NUMBER: 480.76-1(HV)
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: 650-324-1677
18 TELEFAX: 650-324-1678
19 INFORMATION FOR SEQ ID NO: 4:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 5163 base pairs
22 TYPE: nucleic acid
23 STRANDEDNESS: double
24 TOPOLOGY: linear
25 MOLECULE TYPE: DNA (genomic)
26 US-08-928-361B-4

```

Query Match	1.28;	Score 60.6;	DB 3;	Length 5163;
Best Local Similarity	44.28;	Pred. No. 1e-05;		
Matches 249;	Conservative 0;	Mismatches 314;	Indels 0;	Gaps 0;

[illegible]

RESULT 6

QY	4350	TTTAGCAAGAAAGGAATGTGGAGGTTAAATCAAGGAGCTACCTCAAAACAATCA	4409
Db	1034	CTACCAACAAACAAACAACTACTACTACCAACAACAACAACCTACTACTACTA	10933
QY	4410	AGACAAGCTGGAGATTTGCAAGAAAAATAACAATTCGTGGGAATTTGAGAGCTGTAC	4469
Db	1094	CACAAACACACACAAACACACACAACAACACTACACGAAACCAACAACAACAA	11533
QY	4470	CGATTATAACCAACAACATTCG	4492
Db	1154	CAACTACTACTACTACTACACC	1176

RESULT 8

```

: US-08-928-361B-2
: Sequence 2, Application US/08928361B
: Patent No. 6071518
:
: GENERAL INFORMATION:
:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INECTIONS
: NUMBER OF SEQUENCES: 30
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIRSA
: STREET: 385 Sherman Avenue, Suite 6
: City: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: VERNY, HANA
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480.76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 551 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: US-08-928-361B-2

```

Query Match	1.2%;	Score 58.8;	DB 3;	Length 551;
Best Local Similarity	45.2%;	Pred. No. 3.2e-05;		
Matches 216; Conservative	0;	Mismatches 262;	Indels 0;	Gaps 0;

[illegible]

D _b	1214	CTACTACTACTACTACTACCAACAACAACCACAACCAACCACTACCAAGAAACCAA	1273
Oy	4116	GGACTCCATGACACCGGATATCAATTTCGCTAATGTGTGCTGGGGTATTACAGATCTC	4175
D _b	1274	CACCAACCAACCAACAACCAACCAACCACTACTACAACCGACGAACCAACCAACCAAGA	13338
Oy	4176	GAGGAAAAATTACAGCTGCACCTTGACTCTTTAAAAGTATTAACGATAACCAAGG	4235
D _b	1334	CACCAACCAACCAACCACTACTACAACCAACCAACCACTACTACTACTACTACCAACA	13938
Oy	4236	CGAAGATGAAAAATATCTGCCCTTCTCGATTAACATCGAAACCCCTGTACAGACAGTAA	4295
D _b	1394	CACCAACCACTACTACTACTACTACCAACAACAACAACCACTACTACTACTACTACAACCAACA	1453
Oy	4296	CGACAAAAATGACCTCTTGCTAAATTACCTGGAGGCCAAGTCTCCAACATTACTTACGA	4355
D _b	1454	CCACCAACCAACCAACCGAACCCACCACTACCAAGAAACCAACAACAACCAACCACTACTA	1513
Oy	4356	GAACAGCAATGTGAAGTTAAAAATCAAGAGCTGAACCTACTCAAAACCAATCCAAAGC	4413
D _b	1514	CTTACTACTACCAACCAAGAAACCAACCAACAATCACCTACTCTCCCAACAACCAACCTACTAC	1571

RESULT 9

```

US-08-928-361B-1
: Sequence 1, Application US/08928361B
: Patent No. 6071518
:
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INFECTIONS
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Verny, Hana
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480,76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7334 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
US-08-928-361B-1

```

Query Match	1.28;	Score 58.8;	DB 3;	Length 7334;
Best Local Similarity	45.28;	Pred. No. 3.8e-05;		
Matches 216;	Conservative 0;	Mismatches 262;	Indels 0;	Gaps 0;

QY 3205 AAGAGAGCGCGAGATCGCCGAGACAGAGAACACACTCTGAGAAACACCAAGTTCTTCTC 326
 Db 641 TACAAACAACATGGCTACTCTCCAAAAAATCACATATAGTTACTCCCGAGATTACATG 700
 QY 3265 AAMCACTACAAAGCCCTCGTCAAGTATTTATTAATGAGCGACTTCTCTCTGAAGACTCTC 3324
 Db 701 AAGCAACAACAACGGCTTCTCCGAGAGTTACAAACAGAACACACACACACATATTTTC 760
 QY 3325 TCCGAGAGAGACATCCAGACCGAGGATTAACACTACGCCAGCTCGAGAACTTCAGGTCCTG 3384
 Db 761 TCCGAGATTACAAACAACATATTAACAATATGTTTCTCCGAGATTACAAACAACAT 820
 QY 3385 TCTTAAGCTCGAAGCAGCTGGAAGGACACCTAACCTGAGAGAGAAGACTCGACTAC 3444
 Db 821 AATTAACAATATGTTTCTCCGAGATTACAAACAATTAACAACAAATGCTTCTCCGAG 880
 QY 3445 CTCCTTAGCGGACCTGCATCTCACTGATCGCGACCTCAAGGAAGTCATTAAAGAACAGAAC 3504
 Db 881 AACTACAAACAACATATGTTTCTCCGAGATTACAAACAAAAACATTAACAACAATGCT 940
 QY 3505 TATACCGGCAATAGCCCAACGAGAACTTAATACAGAC 3540
 Db 941 TTCTCTGAGATTACAAACAACAAAAACAACAATAGCC 976

```

RESULT 13 452A-24
US-08-276-452A-24
: Sequence 24 Application US/08276452A
: Patent No. 5646029
: GENERAL INFORMATION:
: APPLICANT: Chen, Chao-Guang
: APPLICANT: Mau, Shiao-Lim
: APPLICANT: Du, He
: APPLICANT: Gane, Alison M
: APPLICANT: Bacic, Antony
: APPLICANT: Claire, Adrienne E
: TITLE OF INVENTION: Plant Arabidogalactan Protein (AGP) Genes
: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee and Wimer, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America
: ZIP: 80303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/276,452A
: FILING DATE: 18-JUL-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Caruthers, Jennie M.
: REGISTRATION NUMBER: 34,464
: REFERENCE/DOCKET NUMBER: 27-91A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303)499-8080
: TELEFAX: (303)499-8089
: TELEX: 49617824
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1690 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: '60..1442

```

```

1 FEATURE:
2 NAME/KEY: misc_feature
3 LOCATION: 1..38
4 OTHER INFORMATION: /note="Nucleotide sequence
5 OTHER INFORMATION: obtained by PCR which does not overlap with the
6 OTHER INFORMATION: cDNA clone"
7 FEATURE:
8 NAME/KEY: misc_feature
9 LOCATION: 60..128
10 OTHER INFORMATION: /note="Predicted transmembrane
11 OTHER INFORMATION: segment"
12 FEATURE:
13 NAME/KEY: misc_feature
14 LOCATION: 135..179
15 OTHER INFORMATION: /note="Derived amino acid sequence
16 OTHER INFORMATION: corresponding to the peptide sequence by protein
17 OTHER INFORMATION: microsequencing"
18 FEATURE:
19 NAME/KEY: misc_feature
20 LOCATION: 135..179
21 OTHER INFORMATION: /note="Amino acids 27 to 36, 38,
22 OTHER INFORMATION: and 40 are identical to that in the peptide
23 OTHER INFORMATION: obtained by direct microsequencing"
24 FEATURE:
25 NAME/KEY: misc_feature
26 LOCATION: 135..179
27 OTHER INFORMATION: /note="Amino acid 26 may also be
28 OTHER INFORMATION: Ala; 37 and 39 can also be undetermined residues"
29 OS-08-276-452A-24

```

Query Match	Best Local Similarity	1.0%; Score 48.2; DB 1; Length 1690;
Matches	230; Conservative	0; Mismatches 303; Indels 0; Gaps 0;
QY 859	GACGTCCTCATCTATTAACAACAACAGCTTGAAGAAGCCCATACCTCATCAGCTACTGGAG	918
DB 573	GATGAGTCCTCAATAAACAATGGTTATGATTTCCACAACAACAGACAACAACATGGTAT	632
QY 919	AAGCCATATAGACACCCCTCAGAGAAAGTAAATATCAAAAGAACTGCTGCACAAGATTAT	978
DB 633	GATTCCAACCAATTAACAACAACATATACGATGATGGCTTCTCCGAGAAATTACAACAACAT	692
QY 979	GAATATTAGAATTCCTCCGCCACGCTGGGAACACCCCTTACAGACGCTGCTGGCAAG	1038
DB 693	GGCTACTCGGAAGATGCTAATTAACAATAATACATATGGCTACTCAGAGATTACAACAAC	752
QY 1039	AACAAGAAGATAGAGAGACGACGAGAAAGATCAAGAAGATGCCAAAACCTTAAGTTC	1098
DB 753	AATAACAACATGGCTACGCCAAGATTACAACAATGGCTACTCTCAGATTACAACAAC	812
QY 1099	AACATAGATTCTCTCTTACTGATGATCCCTTGAGTCTGAGTACTTGTGAGAGAGAANT	1158
DB 813	AACATTAATTTTCTACTCGSAGATTTACAACAACAACAACAATGTTTCTCGGGAANT	872
QY 1159	AAGAATATAGACATCTCCGCCAAGCTGAGACAAGAAAGATCAACCGAATATGATAT	1218
DB 873	TCCAACAACATATGGCTACTCCAAAAGATCAACATATATGGCTACTCCGAAATTAACAG	932
QY 1219	CCCAATGGTGTGAGCTACCCCTCTGTCTTATACGATATCAACAAGCTCTCAACGAGCTC	1278
DB 933	AACAACAACAATGGCTTCTCCGAGAGTTTACAACAACAACAACAACAACAACAAC	992
QY 1279	AATAGCTTCGGGAGCTTGATTAAACCCCTTCGATTTATAGAAAGAACCCCTTAAGATATC	1338
DB 993	AACGTTTCTCTGAGAAATTACACAACAACAATTAACATTAATGTTTCTCGAGAAATTAC	1052
QY 1339	TACACAGACAATGAGAGAAAGTTTATCAACGAATTCAGAGAGAGATCAA	1391
DB 1053	AACAACAACATTAACAACAATGCTTTTACGAGAAATTACAACAACAACAACA	1105

```

; Sequence 24, Application US/08798744
; Patent No. 5830747
; GENERAL INFORMATION:
; APPLICANT: Chen, Chao-Guang
; APPLICANT: Mau, Shiao-Lim
; APPLICANT: Du, He
; APPLICANT: Bacic, Antony M
; APPLICANT: Clarke, Adrienne E
; TITLE OF INVENTION: Plant Arabidogalactan Protein (AGP) Genes
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,744
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/276,452
; FILING DATE: 18-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Caruthers, Jennie M.
; REGISTRATION NUMBER: 34,464
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)499-8080
; TELEFAX: (303)499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1690 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 60..1442
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..38
; OTHER INFORMATION: /note="Nucleotide sequence
; OTHER INFORMATION: obtained by PCR which does not overlap with the
; OTHER INFORMATION: cDNA clone"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 60..128
; OTHER INFORMATION: /note="Predicted transmembrane
; OTHER INFORMATION: segment"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 135..179
; OTHER INFORMATION: /note="Derived amino acid sequence
; OTHER INFORMATION: corresponding to the peptide sequence by protein
; OTHER INFORMATION: microsequencing"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 135..179
; OTHER INFORMATION: /note="amino acids 27 to 36, 38,
; OTHER INFORMATION: and 40 are identical to that in the peptide
; OTHER INFORMATION: obtained by direct microsequencing"
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: 135..179
; OTHER INFORMATION: /note="Amino acid 26 may also be
; OTHER INFORMATION: Ala: 37 and 39 can also be undetermined residues"
; US-08-798-744-24
;
; Query Match
; Best local Similarity 43.2%; Pred. No. 0.01;
; Matches 230; Conservative 0; Mismatches 303; Indels 0; Gaps 0;
;
; QY 859 GACCTGTCATCTATTAACAACAGCTTGGAAGAACCCATACCTATCAGCGTACTGGAG 918
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 573 GATGAGTCTCTACATTAACAATGGTATGATTCACAACAACGACAACAACATGTTAT 632
;
; QY 919 AACGGATAGACACCCCTCAAGAGAAATGAAAATCAAGAACTGCTCGCAAGATTAAT 978
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 633 GATTCCAACATATACACAACAAATACGATGAGTCTCTCGAAGATTACACAACAAT 692
;
; QY 979 GAATTAAGAAATCCTCCGCGCCCACTCGGAGACACCCCTAACACGCTGCTGACAAAG 1038
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 693 GCGTACTCGAGATGCTATATACAAAATTAACAATGGCTACACAGAAATTAACAACAAC 752
;
; QY 1039 AACAGAGATAGAGAGACGACGAGAAAGAGATCAAGAGATCGCCAAACCAATTAAGTTC 1098
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 753 AATAACAACAATGGCTACGCGCAAGAAATTACACAATGGCTACTCTCAGAGATTACAACAAC 812
;
; QY 1099 AACATGATTCCTCTTTACTGATCCCTTGAGCTGAGTACTGAGAGAGAAAGAAAT 1158
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 813 AACAAATAATTTTACTCGGAAATTTACACAACAACAACAACAATGTTTTCTCGGAAAT 872
;
; QY 1159 AAGAATATAGACATCTCCGCCAAAGTGCAGACAAAGGAATCAACCGAATTAAGATAT 1218
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 873 TCCAAACAACATGGCTACTCTCCAAAAGATCAACAATAATGGTATGCTCCGAAATTAACATG 932
;
; QY 1219 CCCAATGGTGTGACGTACCTCTGTCTTATTAACGATATCAACAAGCTCTCAACGAGCTC 1278
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 933 AACCAACAACAATGGCTTCTCCGAGATTACACAACAACAACAACAACAACAACAAC 992
;
; QY 1279 AATAGCTCGGTGACTGATTATACCCCTTGATTAACGAAGAAGCCCTGAAGATATTC 1338
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 993 AACGTTTCTCTGAGATTTCACAACAACAACAATTAATATTTTCTCGGAAATTAAC 1052
;
; QY 1339 TACACAGACAATGAGAGAAAGAAAGTTATCAAGAAATCAAGAGAAAGATCAAA 1391
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1053 AACCAACAACAATTAACAACAATGCTTCTACGAAATTAACAACAACAACAACA 1105
;
; RESULT 15
; PCT-US95-10668-3
; Sequence 3, Application PC/TUS9510668
; GENERAL INFORMATION:
; APPLICANT: James Eberwine
; TITLE OF INVENTION: A Method of Sequencing Proteins by
; TITLE OF INVENTION: Epitope Ordering and Protein
; TITLE OF INVENTION: Restriction Mapping
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 210 Lake Drive East, Suite 201
; CITY: Cherry Hill
; STATE: NJ
; COUNTRY: USA
; ZIP: 08002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10668
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/294,133

```

FILED DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
PCT-US95-10668-3

Query Match 0.9%; Score 46.4; DB 5; Length 198;
Best Local Similarity 52.6%; Pred. No. 0.0085;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1357 AAGAGTTTATCAAGCAATCAAGGAGATCAAAATTGAGAGAGAAATTTGAGAGT 1416
DB 2 AAC 61
QY 1417 GACAGAAAAGTTACGAGACGACGACAAAGTCTAAAGATATCACTAAAGATATGAA 1476
DB 62 AAC 121
QY 1477 AAGCTGCTCAAGCAGATCTATGATTCCAAATTCACAAATTAACATGACCTGACCACTTC 1536
DB 122 AAC 181
QY 1537 GAGAAAATGATG 1548
DB 182 AAGCTGATCTG 193

RESULT 16
PCT-US95-10668-4
Sequence 4, Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
PCT-US95-10668-4

Query Match 0.9%; Score 46.4; DB 5; Length 198;
Best Local Similarity 52.6%; Pred. No. 0.0085;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1357 AAGAGTTTATCAAGCAATCAAGGAGATCAAAATTGAGAGAGAAATTTGAGAGT 1416
DB 2 AAC 61
QY 1417 GACAGAAAAGTTACGAGACGACGACAAAGTCTAAAGATATCACTAAAGATATGAA 1476
DB 62 AAC 121
QY 1477 AAGCTGCTCAAGCAGATCTATGATTCCAAATTCACAAATTAACATGACCTGACCACTTC 1536
DB 122 AAC 181
QY 1537 GAGAAAATGATG 1548
DB 182 AAGCTGATCTG 193

RESULT 17
US-08-098-327E-37
Sequence 37, Application US/08098327E
Patent No. 6270771
GENERAL INFORMATION:
APPLICANT: GUBERTIN-MARCHAND, Claudine
TITLE OF INVENTION: DRUGS, PLEI
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPTIOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-34

Query Match 0.9%; Score 46.4; DB 4; Length 988;
Best Local Similarity 45.7%; Pred. No. 0.022;
Matches 161; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 629 TGAAGATCAGAGCCAGGAGTTGAGCTATTTGAAGAGTTGGCTTCGATATCGCAGC 688
DB 596 TAAAGAAAGTTGCAAGACACAAAGCGATTTAGAACAGAGAGAGCTGCTAAAGAAA 655
QY 689 CTCCTGCACACATCAAGACATGTGGAAAGATGGAAGATTATTAATAAAGAAATAGA 748
DB 656 GTTGCAAGAAACAAACGCGATTAGAACAGAGAGAGCTGCTAAAGAAAGTTGCAAGA 715
QY 749 AGACCATCGAAGCATTAAAGAGCTGATGCAAGATCCAAAAGACCATTAGCAAAAATA 808
DB 716 ACAACAAAGCGATTGACAGACAGAGAGCTGCTAAAGAAAGTTGCAAGAACAAAG 775
QY 809 AGATGCAACCAAGAGAGAAAGAAAGAGTTGTACCGCCCGAGTACGACCTGTCCA 868
DB 776 CGATTAGAACAGAGAGAGAGCTGCTAAAGAAAGTTGCAAGAACAAAGCGATTAGA 835
QY 869 TCTATTAACAACAGCTTGAAGAGCCCATTAACCTCATACGCTACTGAGAGAGCGCATAG 928
DB 836 ACAAGAGAGAGCTGCTAAAGAAAGTTAGAACAGAGAGAGCGATTTAGAACAGAGTAG 895
QY 929 ACACCTCAAGAGAAATGAATAATATCAAGAGACTGCTCGACAGATTATGA 980
DB 896 ACTTGCTAAAGAAAGTTGCAAGACAAAGCGATTGAGAACAGAGAGA 947

RESULT 20
US-08-462-625-34
Sequence 34, Application US/08462625
Patent No. 6319502
GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUIHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPTIOPES CAPABLE
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P. O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-462-625-34

Query Match 0.9%; Score 46.4; DB 4; Length 988;
Best Local Similarity 45.7%; Pred. No. 0.022;
Matches 161; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 629 TGAAGATCAGAGCCAGGAGTTGAGCTATTTGAAGAGTTGGCTTCGATATCGCAGC 688
DB 596 TAAAGAAAGTTGCAAGACACAAAGCGATTTAGAACAGAGAGAGCTGCTAAAGAAA 655
QY 689 CTCCTGCACACATCAAGACATGTGGAAAGATGGAAGATTATTAATAAAGAAATAGA 748
DB 656 GTTGCAAGAAACAAACGCGATTAGAACAGAGAGAGCTGCTAAAGAAAGTTGCAAGA 715
QY 749 AGACCATCGAAGCATTAAAGAGCTGATGCAAGATCCAAAAGACCATTAGCAAAAATA 808
DB 716 ACAACAAAGCGATTGACAGACAGAGAGCTGCTAAAGAAAGTTGCAAGAACAAAG 775
QY 809 AGATGCAACCAAGAGAGAAAGAAAGAGTTGTACCGCCCGAGTACGACCTGTCCA 868
DB 776 CGATTAGAACAGAGAGAGAGCTGCTAAAGAAAGTTGCAAGAACAAAGCGATTAGA 835
QY 869 TCTATTAACAACAGCTTGAAGAGCCCATTAACCTCATACGCTACTGAGAGAGCGCATAG 928
DB 836 ACAAGAGAGAGCTGCTAAAGAAAGTTAGAACAGAGAGAGCGATTTAGAACAGAGTAG 895
QY 929 ACACCTCAAGAGAAATGAATAATATCAAGAGACTGCTCGACAGATTATGA 980
DB 896 ACTTGCTAAAGAAAGTTGCAAGACAAAGCGATTGAGAACAGAGAGA 947

RESULT 21
US-08-182-175A-104
Sequence 104, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage proteins with defined structure Containin
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:

COMPUTER READABLE FORM:

Db 242 AAAGATGAAGCAATGAGAAACAAATGAAGTGTGAGAGAGAAATGAAGAGCTGCA 301
QY 3213 AGCCGAGAT 3221
Db 302 AGAGAAAGAT 310

RESULT 23
5171843-10
Patent No. 5171843
APPLICANT: NUSSENZWEIG, VICTOR
TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
PURIFYING IT
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/175,112
FILING DATE: 30-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 754,645
FILING DATE: 9-JUL-1985
APPLICATION NUMBER: 115,634
FILING DATE: 26-OCT-1987
APPLICATION NUMBER: 649,903
FILING DATE: 12-SEP-1984
SEQ ID NO:10
LENGTH: 697
5171843-10

Query Match 0.9%; Score 45.6; DB 6; Length 697;
Best Local Similarity 47.0%; Pred. No. 0.03;
Matches 141; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 1689 GGAGAAAGCAATTAAGTATTACAAAGATTCATAGTAGATCGAAAACGAGATCGAGAC 1748
Db 124 GGACATAGACACTGACACACCGACGAGATAGACATGACACGCCACGAGGAGAC 183
QY 1749 GCTTGTGAGACATTAAAGAGATGAAGACAGTTGTTGAGAAAGATTTACAAAGA 1808
Db 184 AGACGAGATGACACCGACGAGAGACAGATGACAGATGACACCGACGAGAGACGA 243
QY 1809 CGAAATTAACCAAGATGAGAGATCCCTGAGAGTCTCCGATATTGTTAAAGTCCAGTGA 1868
Db 244 GATGACCAACCCAGCTGATAGACGCTGACCAACCCAGCTGATAGACGAGCTGGA 303
QY 1869 GAAGTGCTCCTGATGAACAAGATTGATGAATCAAGACCTCACTTCATTTGAAGAA 1928
Db 304 CAACCGACGAGATGAGACAGATGACACCGACGAGAGATGAGACGCTGACACGCCA 363
QY 1929 CGTGAGATTAAACATATATATACATGTGCCGAATAGTTATAGCAGAGAGATAAGCAGA 1988
Db 364 GCAGGAGATGAGCAGATGACACCGCAGAGGAGATGAGACGCTGACACGCCAGCAGA 423

RESULT 24
5171843-8
Patent No. 5171843
APPLICANT: NUSSENZWEIG, VICTOR
TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
PURIFYING IT
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/175,112
FILING DATE: 30-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 754,645
FILING DATE: 9-JUL-1985
APPLICATION NUMBER: 115,634
FILING DATE: 26-OCT-1987
APPLICATION NUMBER: 649,903
FILING DATE: 12-SEP-1984
SEQ ID NO:8
LENGTH: 1137

5171843-8
Query Match 0.9%; Score 45.6; DB 6; Length 1137;
Best Local Similarity 47.0%; Pred. No. 0.04;
Matches 141; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 1689 GGAGAAAGCAATTAAGTATTACAAAGATTCATAGTAGATCGAAAACGAGATCGAGAC 1748
Db 316 GGAGACAGACGACAGATGACACCGACGAGTGTATGACAGATGACACCGACGAGGAGAT 375
QY 1749 GCTTGTGAGACATTAAAGAGATGAAGACAGTTGTTGAGAAAGATTTACAAAGA 1808
Db 376 AGACGAGCTGACACCGACGAGAGATGACAGATGACACCGACGAGAGACGAGACGA 435
QY 1809 CGAAATTAACCAAGATGAGAGATCCCTGAGAGTCTCCGATATTGTTAAAGTCCAGTGA 1868
Db 436 GATGACACCGCAGACGAGAGACAGATGACACCGACGAGAGACGAGAGATGAGCA 495
QY 1869 GAAGTGCTCCTGATGAACAAGATTGATGAATCAAGACCTCACTTCATTTGAAGAA 1928
Db 496 CAACCGACGAGTATAGCAGCTGACCAACCGACGAGTATGAGCAGCTGACCAACCA 555
QY 1929 CGTGAGATTAAACATATATATACATGTGCCGAATAGTTATAGCAGAGAGATAAGCAGA 1988
Db 556 GCAGGAGATGAGCAGATGACACCGCAGAGGAGATGAGCAGCTGACACGCCAGCAGA 615

RESULT 25
PCT-US95-10668-1
Sequence 1, Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
RESTRICTION Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: NO
PCT-US95-10668-1

Query Match 0.9%; Score 45.4; DB 5; Length 198;
Best Local Similarity 53.0%; Pred. No. 0.016;

ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 950 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-32

Query Match 0.9%; Score 44.8; DB 4; Length 950;
Best Local Similarity 45.5%; Pred. No. 0.058;
Matches 160; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 629 TGAAGATCAGACCAAGCGATTGGACGATTTGAAGAAGTTGGCTTCGATATCGCAAGC 688
DB 29 TAAAGAAAAGTTGCAAGAACCAAGCGATTGAGAACAAAGATAGACTGCTAAAGAAAA 88
QY 689 CTCCTGACACATCAAGACGATGTGGAAAGATGGAAGATTAATTAAGAAATTAAGA 748
DB 89 GTTACAAAGCGACCAAGCGATTGTAACAAGAGAGACTTGTAAAGAAAAGTTGCAAGA 148
QY 749 AGACCATCGAGAACATTTAAGAGCTGATGGAAGATCCAAAAAGACCATAGACAAAAA 808
DB 149 ACAACAAAGCGATCTAGAACAGAGAGACGTCTAAAGAAAAGTTGCAAGAACCAAGA 208
QY 809 ACAAATGCAACCAAGAGAGAAAGAAAGAAAGTGTACACAGGCCCGCTAGCACTGTCCA 868
DB 209 CGATTTAGAACAAGAGAGCGTGTAAAGAAAAGTTGCAAGAACCAAGCGATTTAGA 268
QY 869 TCTATTAACAACAGCTTGAAGAGCCCATACCTCATCGAGCTAGCGAAGCGCATG 928
DB 269 ACAAGATAGACTTGTCTAAAGAAAAGTTACAGAGCGACGACGATTTTGAACAAAGAG 328
QY 929 ACACCCCTCAAGAAGAAATATCAAGAAAGTGTCTGCAAGATTAATGA 980
DB 329 ACGTGCTTAAGAAAAGTTGCAAGAACCAAGCGATTGTAAGAACAGAGAGA 380

RESULT 29
US-08-462-625-32
Sequence 32, Application US/08462625
Patent No. 6319502
GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre

TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 950 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-462-625-32

Query Match 0.9%; Score 44.8; DB 4; Length 950;
Best Local Similarity 45.5%; Pred. No. 0.058;
Matches 160; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 629 TGAAGATCAGACCAAGCGATTGGACGATTTGAAGAAGTTGGCTTCGATATCGCAAGC 688
DB 29 TAAAGAAAAGTTGCAAGAACCAAGCGATTGAGAACAAAGATAGACTGCTAAAGAAAA 88
QY 689 CTCCTGACACATCAAGACGATGTGGAAAGATGGAAGATTAATTAAGAAATTAAGA 748
DB 89 GTTACAAAGCGACCAAGCGATTGTAACAAGAGAGACTTGTAAAGAAAAGTTGCAAGA 148
QY 809 ACAAATGCAACCAAGAGAGAAAGAAAGAAAGTGTACACAGGCCCGCTAGCACTGTCCA 868
DB 209 CGATTTAGAACAAGAGAGCGTGTAAAGAAAAGTTGCAAGAACCAAGCGATTTAGA 268
QY 869 TCTATTAACAACAGCTTGAAGAGCCCATACCTCATCGAGCTAGCGAAGCGCATG 928
DB 269 ACAAGATAGACTTGTCTAAAGAAAAGTTACAGAGCGACGACGATTTTGAACAAAGAG 328
QY 929 ACACCCCTCAAGAAGAAATATCAAGAAAGTGTCTGCAAGATTAATGA 980
DB 329 ACGTGCTTAAGAAAAGTTGCAAGAACCAAGCGATTGTAAGAACAGAGAGA 380

```
RESULT 30
US-08-098-327E-41
: Sequence 41, Application US/08098327E
: Patent No. 6270771
: GENERAL INFORMATION:
: APPLICANT: GUERIN-MARCHAND, Claudine
: APPLICANT: DROULHE, Pierre
: TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
: TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
: TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/098,327E
: FILING DATE: 24-NOV-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 01286
: FILING DATE: 05-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McGowan, Malcolm K.
: REGISTRATION NUMBER: 39,300
: REFERENCE/DOCKET NUMBER: 010830-045
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1482 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1482
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 92/13884
: PUBLICATION DATE: 20-AUG-1992
: US-08-098-327E-41

Query Match          0.9%; Score 44.8; DB 4; Length 1482;
Best Local Similarity 45.5%; Pred. No. 0.075;
Matches 160; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
```

```
QY 869 TCTATACCAACAGCTTGAAGAGCCATTAACCTTCAGCGTACTGAGAGCGCATAG 928
DB 279 ACAAGATGACTCTGCTAAGAAAGTTACAGAGAGCAAGCCATTGAAACAAGAG 338
QY 929 ACACCTCAGAGAGATGAATATCAAGAAGCTGCTGACAGATTATGA 980
DB 339 ACGTCTAAGAAAGTTGCAAGAACAAAGCATTTAGAACAGAGAGA 390

RESULT 31
US-08-098-327E-45
: Sequence 45, Application US/08098327E
: Patent No. 6270771
: GENERAL INFORMATION:
: APPLICANT: GUERIN-MARCHAND, Claudine
: APPLICANT: DROULHE, Pierre
: TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
: TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
: TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/098,327E
: FILING DATE: 24-NOV-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 01286
: FILING DATE: 05-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McGowan, Malcolm K.
: REGISTRATION NUMBER: 39,300
: REFERENCE/DOCKET NUMBER: 010830-045
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 45:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1482 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1482
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 92/13884
: PUBLICATION DATE: 20-AUG-1992
: US-08-098-327E-45

Query Match          0.9%; Score 44.8; DB 4; Length 1482;
Best Local Similarity 45.5%; Pred. No. 0.075;
Matches 160; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
```

```
Query Match      0.98; Score 44.8; DB 4; Length 1482;
```

	Best Local Similarity	45.5%	Pred. No. 0.075;	
	Matches 160;	Conservative	0;	Mismatches 192; Indels 0; Gaps 0;
QY	623	TGAACATCAGACCACACGATGGTGGACGTATTGGAGAAGTTGGTCTTGGCATATCGCAAGC	688	
Db	39	TAAACAAAAAGTTGGCAAGAACCAACAAAGGATTTAAACAAAGATAGACTTGGCTAAAGAAAA	98	
QY	689	CTCTGGACACATCTCAGAGACAACTGTGGGAAAGATGGAGATTTATTTTAAAAAAGTAAGA	748	
Db	99	GTTACAAAGGCGCAACGCAATTTAGAACCAAGAGAGACTTGGTAAAGAAAAAGTTGGCAAGA	158	
QY	749	AGACCATGAGACACTTTAAAGCGTCTGATCGAAGATCCAAAAAGACCCTAGACAAAAATA	808	
Db	159	ACAAACAAAGCGATCTTACAGAACCAAGAGAGACGGTCTTAAAGAAAAAGTTGCAAGAACCAANAAG	218	
QY	809	AGAAATCGCAACCAAGAGAGAGAAAGAAAGAAAGTTGTACCAAGGCGCAGACACACTGTGCA	868	
Db	219	CGATTTAGACACACAGAGACGCTGCTAAAGAAAAAGTTGCACAGAACCAACCAAGCATTTAGA	278	
QY	869	TCTATTAACAAACAGCTTGAAGAGAGCCCATTAACCTTCATCAGCGTACTGAGAAAGCGCATAG	928	
Db	279	ACAAGATAGACTCTGCTTAAAGAAAGTTTACAAAGAGACAGCAAAAGCCGATTTAGAACCAAGAGAG	338	
QY	929	ACACCCCTGAAGAGAAATGAATAATATCAAGAAAGCTGCTGACAGAGATTAATGA	980	
Db	339	ACGTGCTTAAGAAAAAGTTGCAGAACACAAAGCCATTTTAGACACAGAGAGAC	390	

? Sequence 45 Application US/08462625
? Patent No. 6319502
? GENERAL INFORMATION:
? APPLICANT: GUERIN-MARCHAND, Claudine
? APPLICANT: DRUILHE, Pierre
? TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
? TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
? TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
? NUMBER OF SEQUENCES: 46
? CORRESPONDENCE ADDRESS: |
? ADDRESSEE: |

ADDRESS: E. Burns, Dodan, Swecker & MacInis
STREET: P. O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA: 1
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 1482 base pairs
TYPE: nucleic acid
STRANDEDNESS: single


```

; SEQUENCE CHARACTERISTIC
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
US-09-098-487-2

```

Query Match	0.9%	Score	44.2	DB	2	Length	2277
Best Local Similarity	26.7%	Pred	No. 0.14				
Matches	185	Conservative	102	Mismatches	402	Indels	3
						Gaps	1

OY	1625	ATCTTGGAGAGCTACCAAGACCTCTTAAGATATGTGGAGCATCTCTCGCGAACATTG	1684
Db	170	ATGTCNCAARMGNCGYTAAAGARATACACNANTTTCARBGNITGYAARATHAAYGTNCAG	229
OY	1685	TTTGTGAGAAAGAACTAAAGTATTACAGAAITCTCTAAGTAGACTCGAAAACGAGATCG	1744
Db	230	TGNCNNAARAARAAYTTWNGNAAAYAAACNAAACGARAARBGNAARAAYGARAAYWNGART	289
OY	1745	AGACGCTGTGTGACAACATTTAAGAGAGTGAACAACAGTGTGTTTGAGAAAGAGATTACA	1804
Db	230	GYCCNNAARAARAGACCNAARGCNNAARARCNARCTNGCNGATARAARCGCMNNTYNA	349
OY	1805	AAGACGAAAAATAAACGAGTAGAGAGATCCTGGAGTCTCCGATATTTGTTAAAGTCAC	1864
Db	350	THATHMGMAAYTTWMSNTTYAARTGYWNSNGARGAYGAYTTNABACNGNTTNYGCNCART	409
OY	1865	TGCGAAGAGGTCTCTCTCATGTACACAGATTGTATCACTCAAGAAGACTCAACTCATTTG	1924
Db	410	TYGGNGCNGTNTYTGGARTAAATATCCNMGNNAARCNAGYGGNABARATGMNGGNTTYG	469
OY	1925	AGAAGCTGGAAGTTTAAACATATATACATGTCCGCAATAGTTATAGCGAGAAATAAGC	1984
Db	470	GNTTYGTNCARTTYTAARAAYTYNTYTNGARBCNGCNGNABRCNTTNAARSGNATGAAYATGA	529
OY	1985	AGCAACCATCTACTACCTCA---TCGTACTCAAGAAAGATAGACAAACTGAAAAGTTTCA	2041
Db	530	ARGARATHAAGGWMGNACNGTNGCNGTNGATYGGCNGTNGCNAARGAARYARTTYAARG	589
OY	2042	TGCCCAAAGTCGAGAGCCCTATCAACGAGAGAAGAAACATTAAACTGGAAGACAGT	2101
Db	590	AYACNCARMSNGTWMNSGCAATHGNCARBARAARBSNCAIYGRMSNMAARCAIYCARGARW	649
OY	2102	CAGATTAACCTCCGAGCTTCCACAGAAAGAGAGATTAACGGACAGCTTACCACCAAGCCCG	2161
Db	650	SNGTNNAARAARAAGGWMGNAGBARGARBARATYTGARBARGARBARAAYGAYGAYG	709
OY	2162	GACAACAGGCGCGTTCAGCTCTGGAAGGGCATAGCGCTGCAAGCTCAAGCAAGAGACGA	2221
Db	710	ATG	769
OY	2222	AGCAGGACACACCTCCAGTCCAGTCCAGTCCAGTCCAGAGGCTAAAGCTCAAGTGCCTACAC	2281
Db	770	THGARMSNAARGTNACNNAARCNNGTNCARATHICARAARBMGNGCNGTNAARBMGNCNGNC	829
OY	2282	CACCAAGCTCCTGTGATTAACAACAACGAGAAAT	2313
Db	830	CNGCNAARWMSWMSNGATCAATWMSGARBARAY	861

RESULT 39
US-08-559-896B-1
Sequence 1, Application US/08559896B
Patent No. 6310046
GENERAL INFORMATION:
APPLICANT: Patrick E. Duffy
APPLICANT: Christian F. Ockenhouse
TITLE OF INVENTION: SEQUESTIRIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: John Moran
STREET: USA MRC - MCMR-JA

CITY: FORT DETRICK, FREDERICK
 STATE: MARYLAND
 COUNTRY: USA
 ZIP: 21702-5012
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: Apple Macintosh
 OPERATING SYSTEM: Macintosh 7.5
 SOFTWARE: Microsoft Word
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/559,896B
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Moran, John
 REGISTRATION NUMBER: 26,313
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 619-2065
 TELEFAX: (301) 619-7714
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1956 base pairs
 TYPE: Nucleic acid
 STRANDEDNESS: Double
 TOPOLOGY: Linear

US-08-559-896B-1

Query Match	0.98;	Score 43.6;	DB 4;	Length 1956;
Best Local Similarity	42.9%;	Pred. No. 0.19;		
Matches 217;	Conservative	0;	Mismatches 289;	Indels 0;
				Gaps 0.

QY	1360	AAATTTTCACGAATTC	CAAGGGAAGATTC	CAAAATTC	GACAAACAAACAAATTC	GAGGTAC	1419
Db	1132	AAATTTAAGCGAAATTC	AAAAGATTAATTTAT	ATTCATAGAAATTT	AAAAATTAATTAATCT	CAC	1191
QY	1420	AGGAAAGTTACGACG	AGCCGACCAAAAGCT	CTAAACGATATC	CTAAAGAGATG	AAAG	1479
Db	1192	CAAAAAAATTAAGAA	AGAAATGTAAATGT	ATTTATATACGAC	ACATTAATGATGT	AT	1251
QY	1480	CTGCTGACGAGATCT	ATGATTC	CAAAATTC	CAACATATACATCG	ACCTGACCACTTGAG	1539
Db	1252	AATATATATATATAT	ATATATAGAGTGT	GAATTAATTTAAATTAAT	TACGACACAAATAT		1311
QY	1540	AAAAATGATGGAAAA	CGGTACTCTTCA	AAAGTGGAGAACT	CACACACCATTATACCTT		1599
Db	1312	AATTTATGAAAAATG	ATGAAAGTTGAATGT	ATACGAAATTTAG	ACAAAGATTAAG		1371
QY	1600	GCATCTATGAGAAATC	TATAGCATATCTTG	AGAACCTC	CAACAAAGCTCTT	TAATATG	1659
Db	1372	GCCAGATAGAGATTA	TATATAGCTATTTT	TACAAAGAAATTT	AAAAAGACAAAAATGT		1431
QY	1660	GAGGACATTTCTCG	CGGAATTTGTGTG	AGAAAGAACTAAAGTAT	TACAGAAATCTC		1719
Db	1432	AATGTTCCATATAT	AGTGAATTTT	TAAATTC	AAAGTGAAGAAAGATTA	CAACACCAATT	1491
QY	1720	ATAAGTAAATCGAAA	CGAGATCGAG	CGCTGTTG	GAGACATTTA	GAGAGATGA	1779
Db	1492	CAACATTAAGAAAG	AAATCAAGTAGATGT	TCAGGAAAAATATTC	CAGTTATTC	AAAG	1551
QY	1780	CAGTTGTTTGAGAG	AGATTTAC	AAAAAGACG	AAATTAACCCGATG	GAGAAAGTCTCGAG	1839
Db	1552	GATTAATTTAAAAA	ATAAGGCCCA	AAAGATTAAC	ACTGAAATGTT	AGATTAATTAAGGA	1611
QY	1840	GTCCTCGATATGT	TGAATGCCAGT				1865
Db	1612	ATAACAATATGTAT	TAATAAAAAATGT				1637

RESULT 4C

[illegible]

RESULT 41
 US-08-471-033-42
 Sequence 42, Application US/08471033
 Patent No. 3770696
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Koziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalin M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,033
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHEICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "Maize optimized DNA"
OTHER INFORMATION: sequence encoding VIP2a(a) with the Bacillus secretion signal
OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
OTHER INFORMATION: contained in pcIB5528"
US-08-471-033-42

Query Match Best Local Similarity 44.7%; Score 43; DB 1; Length 1241;

Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

OY 2972 AGAAGCTGGAAGGAGCATATTAAGCTGAAGAAGACACTGCACTGACCTGT 3031
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 88 AGAAGGAGGAGGAGTGAAGCTTACCCGACCGAGAGGCGCAAGATGAACAATCTCTGG 147
OY 3032 ACAACAAGTCAAACTGAAGTGAAGAGACTCTTCGACAAAGAAGACAGTGGCAGT 3091
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 148 ACAACAAGGAGCATTAAGACCACTACCAAGGATCACCCTTACGATAGCCGGCAGCT 207
OY 3092 ATAGATGCAGATCAGAAAGTTCCTGCTCAAGAGAGAGCTTGA-----AAGCAAC 3145
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 208 TCGAGGAGGAGATCAAGAGCTGCAAGAGATCGACAAGATGTTCGACAAGACCAACTGA 267
OY 3146 TCAACTACGGAACATCCGAACACTGACGAACTTCTGAGTGTCTTCAACAAGA 3205
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 268 GCAACAGCATATCATCTTACCAAGAGAGTGAAGCCACCACCTGCTTCAACAAGAGCC 327
OY 3206 AGAAGGAGGCGGAGATGCGGAGACAGAGAAGAACTGTGAGAACCAAGATTCTTCA 3265
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 328 TGACCGAGGCGCAACCATCAACGAGCGACCGCATGGCCAGTTCAAGAGACAGTCTGG 387
OY 3266 AACACTACAAGGCGCTGCTAAGTATTATTAATGCGAGTCTTCTCTGTAAGACTCTT 3325
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 388 ACCGCGACATCAAGTTGAGAGTCACTGACACCCACTGACCGCCAGAGGTGAGCA 447
OY 3326 CCGAGGAGCATCCACGAGGATACATACGCGGACCTCGAGAACTTCAAGTCTCTGT 3385
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 448 GCAAGGAGCGCGTATCTTAAGGTGACCGTCCCGAGGCGAAGGAGCAACCCCA 507
OY 3386 CTAGCTCGAAGGCAAGCTGAAGGCAACCTGAGAGAAAGAAAGTCACTAC 3445
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 508 CCAAGCGCGCGTATCTTGAACAACAGCGAGTACAAAGATGCTGATGACAGCGGTACA 567
OY 3446 T 3446
DB 568 T 568

APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estuch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHEICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "Maize optimized DNA"
OTHER INFORMATION: sequence encoding VIP2a(a) with the Bacillus secretion sig
OTHER INFORMATION: removed as contained in pcIB5528"
US-08-471-044-39

Query Match Best Local Similarity 44.7%; Score 43; DB 2; Length 1241;

Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

OY 2972 AGAAGCTGGAAGGAGCATATTAAGCTGAAGAAGACACTGCACTGACCTGT 3031
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 88 AGAAGGAGGAGGAGTGAAGCTTACCCGACCGAGAGGCGCAAGATGAACAATCTCTGG 147
OY 3032 ACAACAAGTCAAACTGAAGTGAAGAGACTCTTCGACAAAGAAGACAGTGGCAGT 3091
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 148 ACAACAAGGAGCATTAAGACCACTACCAAGGATCACCCTTACGATAGCCGGCAGCT 207

OY	3092	ATTAAGTATGCAATCAAGAAGTTGACTCTGTCTCAAGSAGCAGCTTGA-----AAGCAAAC	3145
Db	208	TTCGAGCACGGATTCAAGGACTTGAAAGAGATGACAAGATGTTTCGACAAGACCACCTCGA	267
OY	3146	TCAACTCTACTGAACTAACGCCGAAACACGTAAGCTGCAGAACTTCTCAATGTTCTTCAACAAGA	3205
Db	268	GCAACACACATCATTCACCTTACAAAGAACAGTGGAGCCACACCACTGGGCTTCAACAAAGACC	327
OY	3206	AGAAGGAAGCCGAGATGCGCCGAGACAGACAACCTCTGGAGAACCAAGATTCTTCTCA	3265
Db	328	TGACCGGGGCAACACCATCAACAGCGAGCCATGGCCAGTTTCAAGGAGCAGTTCTGG	387
OY	3266	AACACTCAAAAGGCGCTGCTCAACTATTATAATATGGGAGTCTTCTCTGAAAGACTCTCT	3325
Db	388	ACCCCGCATCAAGTTGACTGACGCTTACCTTGACACCCACCTGACCCGCCACGAGAGGTGAGCA	447
OY	3326	CCGAGAGAGACATTCACAGACCAGGATTAATACGCCAGCCTCGAGAACTTCAAGGCTCGT	3385
Db	448	GCAAGGAGCCCGTGATCTGAAAGGTACCCGTCCCACGGCGGCAAGGGCAGCACACCCCCA	507
OY	3386	CTAAGCTCGAAGGCAAGCTGAAGACACCTGAACCTGGAGAGAAAGCTCAAGCTTACC	3445
Db	508	CCAAGCGCGCGTGATCTCTGAACAACAGCAGGAGTACAAGATGCTGATCGCAACAGGCTTACA	567
OY	3446	T	3446
Db	568	T	568

APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spuill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8689
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "maize optimized DNA"
OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion signal
OTHER INFORMATION: removed as contained in PCIB5527
US-08-463-483A-39
Query Match 0.9%; Score 43; DB 2; Length 1241;
Best Local Similarity 44.7%; Pred. No. 0.2;
Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 3146 TCAACTGAGTGAACATTCGAAACACGTACTGCGAAGCTTCTCAGTCTTCTTCACAGAGA 3205
DB 268 GCACAGCATCATCACCCTACAGACAGTGGAGCCACACACATCGGGCTTCACAGAGGCC 327
QY 3206 AGAAGAGAGCCGAGATGCGCGAGACAGACACTCTGGAGAACACCAAGATTCTTCTCA 3265
DB 328 TGACCCAGGGCAACACCATTCACAGGAGCCCATGCGCCCTTCACAGAGGACGACTTCTCT 387
QY 3266 AACACTACAAAGCCCTGCTCAATATTATTAATGCGAGTCTTCTCTGCAAGACTCTCT 3325
DB 388 ACCGACATCAAGTTGACAGAGTACCTGACACCCACCTGACCCGACAGAGTAGCA 447
QY 3326 CCGAGAGAGCATCCAGACCGAGATTAATACGCGACCTCGAGAACTTCAAGTCTCTG 3385
DB 448 GCAAGAGCGCGGTATCTGAAAGTAGTACCGTCCCGAGCGGCAAGGCGACAGCACCCCA 507
QY 3386 CTAAGCTCGAAGCAAGCTGTAAGAGACAACTGAACCTGGAGAGAAAGAAAGTCACTAC 3445
DB 508 CCAAGGCGCGCGGTATCTGAAAGTAGTACCGAGATGATGATGACAAAGGCTTACA 567
QY 3446 T 3446
DB 568 T 568
RESULT 45
US-08-463-483A-42
Sequence 42, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spuill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615


```

Db      268  GCAAGAGCATCATCACCTACAAAGAGCGTGGAGCCCAACACCATCGCTTCAACAGAGCC 327
Oy      3206  AGAAGAGAGCGGATCGGCGAGACAGAGACACTCTGGAGAACACCAATCTTCTCTCA 3265
Db      328  TGACCGAGGCAACACCATCAACAGCGCCATGGCCCGTTCAAGAGCACTTCTCTGG 387
Oy      3266  AACCTACAAAGCGCTCTGCTCAAGTATTAATGCGAGTCTTCTCTGTAAGACTCTCT 3325
Db      388  ACCGCGACATCAAGTTGACAGCTACCTGGACACCACTGACCGCCGACAGGTGAGCA 447
Oy      3326  CCGAGAGACATCCAGACCGAGGATTAACCCAGCTCGAGAACTTCAAGTCTCTGT 3385
Db      448  GCAGAGAGCGCGTGAATCTGAAGTACCGTCCCGAGCGGCAAGGCGCACCCGCCA 507
Oy      3386  CTAAGCTCGAAGCGCAAGTGAAGACACCTGAACCTGGAGAGAAAGACCTGATACC 3445
Db      508  CCAAGGCCGCGGTGATCTGTAACACAGCGAGTACAAAGATGCTGATCGACACGGCTACA 567
Oy      3446  T 3446
Db      568  T 568

```

```

RESULT 47
US-08-471-046A-42
; Sequence 42, Application US/08471046A
; Patent No. 5866326
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5866326artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,046A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ. ID NO.: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1241 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9...1238
; OTHER INFORMATION:
; OTHER INFORMATION: sequence encoding VIP2a(a) with the Bacillus secretion sig
; OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
; OTHER INFORMATION: contained in pCIB538"
; US-08-471-046A-42

```

Query Match

0.9%; Score 43; DB 2; Length 1241;

Best Local Similarity 44.7%; Pred. No. 0.2; Mismatches 260; Indels 6; Gaps 1;

Matches 215; Conservativ

```

Oy      2972  AGAAGCTGAAGAGGACATTAATAGCTGAAGAAGACACTGCACTGAGCTTGACCTGT 3031
Db      88  AGAAGGAGAGAGGTGAGAACTTACCGCCACGAGAGGCGCAAGATGAAACATTCCTGG 147
Oy      3032  ACANCAAGTACAACTGAACACTGGAGAGACTTTCGACAAAGAAGAGAGCGTGGCAAGT 3091
Db      148  ACAACAGAGAGACATCAACACCACTACAAAGAGATCACTTACGATAGCCGGAGCT 207
Oy      3092  ATAAAGTCAGATCAAGAAATGACTCTGCTCAAGAGCAGCTTGA-----AAGCAAC 3145
Db      208  TCAGAGACGAGATCAAGAGACTGAAGAGATCAAGAAAGATGCTTCAAGACCACTGGA 267
Oy      3146  TCAACTCATGAAACAATCCGAACACGCTACTGCAGAACTTCTGAGTGTCTTCAACAGA 3205
Db      268  GCAACGACATCATCACTTAACAAAGAGTGAGGCCACACCATTCGCTTCAACAAAGGCC 327
Oy      3206  AGAAGAGAGCGGATCGCCGAGACAGAGAACTCTGGAGAACCAACCAATCTTCTCTCA 3265
Db      328  TGACCGAGGCAACACCATTAACAGCGCCATGGCCCGTTCAAGAGCACTTCTCTGG 387
Oy      3266  AACCTACAAAGCGCTCTGTAAGTATTAATGCGAGTCTTCTCTGTAAGACTCTCT 3325
Db      388  ACCGCGACATCAAGTTGACAGCTACCTGGACACCACTGACCGCCGACAGGTGAGCA 447
Oy      3326  CCGAGAGACATCCAGACCGAGGATTAACCCAGCTCGAGAACTTCAAGTCTCTGT 3385
Db      448  GCAGAGAGCGCGTGAATCTGAAGTACCGTCCCGAGCGGCAAGGCGCACCCGCCA 507
Oy      3386  CTAAGCTCGAAGCGCAAGCTTAAGAGCAACTGAACCTGGAGAAAGAAAGAACTGATACC 3445
Db      508  CCAAGGCCGCGGTGATCTGTAACACAGCGAGTACAAAGATGCTGATCGACACGGCTACA 567
Oy      3446  T 3446
Db      568  T 568

```

RESULT 48

US-08-470-566B-39

; Sequence 39, Application US/08470566B

; Patent No. 5872212

; GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W

; APPLICANT: Koziel, Michael G

; APPLICANT: Mullins, Martha A

; APPLICANT: Nye, Gordon J

; APPLICANT: Carr, Brian

; APPLICANT: Desai, Nalini M

APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGCI695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal"
US-08-470-566B-39 removed as contained in PCIB5527"

OY	3146	TCACCTCAGCAAGCAATTCGCAAAACAGCTACGTGAGACACTCTTCAGATGTTCTTCAACAGA	3205
Db	268	GCAACAGCATTCATACCTACCAAAAGCTGGAGGCCACACACTGGGGTTTCAACAAGACC	327
OY	3206	AGAAGAAGCCGAGATGCGCGAGACAGAGAACACTCTGGAGAACCAAGAATTCTTCA	3265
Db	328	TGACCGGAGGGCAACCATTAACAAGGAGCCCATGGCCCATTTCAAGNACAGTTCTGG	387
OY	3266	AACACTCAAAAGGCGCTCGTCACAGTATTATTAATGCGAGTCTTCTCCTGGAAGACTCTCT	3325
Db	388	ACCGGACATTCAGTTGACAGCTACCTGGACACCCACTGACGCCCCAGCAGGTGAGACA	447
OY	3326	CCGAGGAGAGCATTCAGACCCGAGGATTACTACGCCACGCTCGAGAATTCAAGGTCTGT	3385
Db	448	GCAAGGAGCGGTGATCTCTGAAAGGATGACCGTCCCGACGGCAAGGGCAACACACCCCA	507
OY	3386	CTAAGCTCGAAGGCAAGCTTAAGGACAACCTGACCTGGAGAGAAGAAGCTCAAGTACC	3445
Db	508	CCAAGGCGCGCTGATCTCTGAAACACAGGAGTACAAAGATGCTGATCGACAAACGCTACA	567
OY	3446	T 3446	
Db	568	T 568	

RESULT 49
 US-08-470-566B-42
 Sequence 42, Application US/08470566B
 Patent No. 5872212
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Kozziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalini M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: No. 5872212artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patencin Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/470,566B
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Melys, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4

Accession	Sequence	Position
Db	GCACACGATCATCTACCTTACCAACAGCTGGAGGCCACGACCATGTGGCTTCAACAAGAAC	327
Oy	AGAAAGAACGCCAGATTCGCCGAGACAGAGAACACTCTGGAGAACACCAAGATTCTTCTCA	32655
Db	TGACCGGAGGCGCAACACCATCAACAGCAGCGCATGGCCCACTTCAAGAGAGCGAGTTCCTGG	387
Oy	AACACTCAAAAGGCGCTCTCACTATTATTAATGCGAGCTTCTCTCTGAAAGACTCTCT	33255
Db	ACCGCGCATCAAGTTCTGACAGCTTACCTTGACACCCACTGACCGCCCAAGCAGGTGAGCA	447
Oy	CCGAGAGAGCATTCAGACCCGAGGATTAACGCCAGCTCGACAGAACTTCAAGTCTCTGT	33855
Db	GCAAGGAGCCCGTGATCTCTGAAGGTACCGTCCCAAGCGGCAAGGGCAGCACACCACTCCA	507
Oy	CTAAGCTCGAAGGCCAAGCTGAAGACAACTGAACTTGAGAAAGAAAGACTCACTTACC	34455
Db	CCAAGCGCGGCGTGATCTCTGAAACACAGGAGTACAAAGTGTGATCGACAACGGCTACA	567
Oy	3446 T 3446	
Db	568 T 568	

RESULT 51
US-08-469-334-42
: Sequence 42, Application US/08469334

```

1  GENERAL INFORMATION:
2
3  APPLICANT: Warren, Gregory W
4
5  APPLICANT: Kozziel, Michael G
6
7  APPLICANT: Mullins, Martha A
8
9  APPLICANT: Nye, Gordon J
10
11 APPLICANT: Carr, Brian
12
13 APPLICANT: Desai, Nalini M
14
15 APPLICANT: Kostichka, N. Kristy
16
17 APPLICANT: Duck, Nicholas B
18
19 APPLICANT: Estruch, Juan J
20
21 TITLE OF INVENTION: No. 5990383e1 Pesticidal Proteins and Strains
22
23 NUMBER OF SEQUENCES: 50
24
25 CORRESPONDENCE ADDRESSES:

```

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: PatentIn Release #1.0, Version #1.30HE
6  CURRENT APPLICATION DATA:

```

```

? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ ID NO: 42:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1241 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Synthetic DNA"
? HYPOTHEetical: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 9..1238
? OTHER INFORMATION:
? OTHER INFORMATION:
? OTHER INFORMATION:
? OTHER INFORMATION:
?
US-08-469-334-42

```

Query Match	0.98;	Score 43;	DB 2;	Length 1241;
Best Local Similarity	44.7%;	Pred. No. 0.2;		
Matches 215; Conservative	0;	Mismatches 260;	Indels 6;	Gaps 1;

Oy	2972	AGAACTCGGAGACGACATCAATTAAGCTGTGAAGAAGCACTGCACACTTCCGACCTGT	30313
Dd	88	AGAAAGGAAAGAGTGGAACTTACCGCCACCGAAGAGGCAAGATGAAACAATCTTCCTGG	147
Oy	3032	ACAACAAGTACCAAACCTGAAACTGAGAGACTTTGACACAAGAAGAACGACGTGGCAGT	3091
Dd	148	ACAACAAGAAACGACATCAAGACCAACTACAGAGAGATCACTTGACATATGCCGGCAGCT	207
Oy	3092	ATAAGATGACGATCAACAAAGTTGACCTGCTCAAGAGAGAGCTTGA-----AACCAAC	3145
Dd	208	TGCGAGGACGAGATTCBAAGCACTTGAGAGAGATGACACAAGATGTTTCGACAAGACCACTCGA	267
Oy	3146	TCAACTCACTGAACAAATCCGAAACAGCTACGTGCGAAGCTTCTCAGTGTTCYCAACAAGA	3205
Dd	268	GCAACACATCATCATCTACTACAAAGACGTGGAGGCCACCACTCGGCTTCAACAAGAGCC	327
Oy	3206	AGAAAGGAAGCGGAGATGGCGGACAGACAGAAACCTCTGSAGAAACCAAGATTCTTCCA	3265
Dd	328	TGACCGGAGGCAACACCATATACACGAGCACGCCATGGCCCCAGTTCAAGAGACAGTTCTCTGG	387
Oy	3266	AACACTACAAAAGGCGCTGTCAGATATTATATATGCGAGCTTCTCTCTCGAAGACTCTCT	3325
Dd	388	ACCCGACATCAAGTTCTGCAGCTACTCTGACACCACTGACCGGCCAGCGAGTGGAGCA	447
Oy	3326	CCGAGAGAGAGCATCCAGACCGAGAGTAACTACCCCACTCGAGAACTTCAAGGTCCTGT	3385
Dd	448	GCAAGGAGCGCGGTATCTTGAAGTAGACGTCCCAAGCGGCAAGGGCAACACCAACCCCA	507
Oy	3386	CTAAGCTCGAAGGCAACGCTCAAGGCAACCTGCAACCTTGAGAGAAGAAGACGCTCACACTAC	3445
Dd	508	CCAAGGCGCGGGGTATCTCTGAACAACAGGAGTACAAAGATGCTGATCGACACAGCGCTACA	567
Oy	3446	T 3446	
Dd	568	T 568	

FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Spullin, W. Murray
 REGISTRATION NUMBER: 32,943
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8615

RESULT 52
 US-09-300-529-39
 ; Sequence 39, Application US/09300529
 ; Patent No. 6066783
 ; GENERAL INFORMATION:
 ; APPLICANT: Warren, Gregory W
 ; APPLICANT: Koziel, Michael G
 ; APPLICANT: Mullins, Martha A
 ; APPLICANT: Nye, Gordon J
 ; APPLICANT: Carr, Brian
 ; APPLICANT: Desai, Nalini M
 ; APPLICANT: Kostichka, N. Kristy
 ; APPLICANT: Duck, Nicholas B

```

1  APPLICANT: Estruch, Juan J
2  TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
3  NUMBER OF SEQUENCES: 50
4  CORRESPONDENCE ADDRESSES:
5  ADDRESSEE: No. 606678artis Corporation
6  STREET: 3054 Cornwallis Road
7  CITY: Research Triangle Park
8  STATE: NC
9  COUNTRY: USA
10 ZIP: 27709
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: floppy disk
14 OPERATOR: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.30B
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/09/300,529
19 FILING DATE: TBA
20 CLASSIFICATION:
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 08/469,334
23 FILING DATE: 06-JUN-1995
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 08/463,483
26 FILING DATE: 05-JUN-1995
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 08/314,594
29 FILING DATE: 09-SEP-1994
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/218,018
32 FILING DATE: 23-MAR-1994
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 08/037,057
35 FILING DATE: 25-MAR-1993
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Meigs, J. Timothy
38 REGISTRATION NUMBER: 38,241
39 REFERENCE/DOCKET NUMBER: S-19506L
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: 919-541-8587
42 TELEFAX: 919-541-8689
43 INFORMATION FOR SEQ ID NO: 39:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 1241 base pairs
46 TYPE: nucleic acid
47 STRANDEDNESS: single
48 TOPOLOGY: linear
49 MOLECULE TYPE: other nucleic acid
50 DESCRIPTION: /desc = "Synthetic DNA"
51
52 HYPOTHETICAL: NO
53
54 FEATURE:
55 NAME/KEY: CDS
56 LOCATION: 9..1238
57 OTHER INFORMATION: /note= "Maize optimized DNA
58 OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal
59 OTHER INFORMATION: removed as contained in pcIB5527"
60
61 US-09-300-529-39

```

Query Match	Similarity	Score	DB	Length
Best	Local	44.7%	Pred.	0.2;
Matches	215;	Conservative	0;	Mismatches
			260;	Indels
				Gaps
				1
Qy	2972	AGAACTGGAAAGACATCATTAAGCTTAAAGACACTGTGAACCTTGACACTGT	3031	
Db	88	AGAAAGAGAGAGAGTGTGGAAAGCTTACCAGCCAGGAGAGGCAAGATGAACTTCTCTG	147	
Qy	3032	ACAACAAGTACAACTGAAACTGGAGAGACTTTCGACAAAGAGACAGTGGCAGAT	3091	
Db	148	ACAACAAGAGACATCAACCAAGCACTACAAAGAGATCACTTCAGCATTAAGCCGGCAGCT	207	
Qy	3092	ATTAAGTGCAGATCAAGAAGTTGACTTCGCTCAAGAGCAGCCTTGA-----AAGCAAC	3145	
Db	208	TTCAGAGACAGAGTCAAGGACCTCTGAAGAGAAATGACAAAGATGTTTCGACAGACCAACCTGA	267	

```

QY 3146 TCACCTACCTCAACAAATCCGAACACGACTGACGAACCTTCACAGTCTTCACACACAGA 3205
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 GCACACGATCATATACCTACACAGACCTGGAGGCCACACACATCGCTTCAACACAGACC 327
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3206 AGAAGGAGCCGACATGCGCCGAGACAGAACACACTCTGGAGAACACACAGATTCCTCA 3265
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 328 TGACCGAGGCGCACACCATCACAGGACGACCATGGCCAGTTCAAGGAGCAGTTCTGG 387
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3266 AACACTACAAAGCCCTGTCACAGTATATATATGGCCAGTCTTCTCTGGAAGACTCTCT 3325
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 ACCGCGACATCAAGTTGACAGAGCTACCTGGACACCCACCTGACCGCCGACGAGTGACGA 447
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3326 CCGAGAGACGATCCACAGCCGAGATTAATACGCCACGCTCGAAGCTTCAGAGTCTGT 3385
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 GCMAAGGACCGGCTATCTCTGAAGTACGCTCCCGACGGCAAGGGCAGACACACCCCA 507
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3386 CTAACTCGAAGGCAAGCTGAAGGACAACTGAACCTGGAGAGAAGAAAGCTCAGCTACG 3445
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 CCAAGCGCGCGGTGATCTGTGAACACGACGAGTACAAAGATGCTGATCGACACAGGCTACA 567
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3446 T 3446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 568 T 568
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 53
US-09-300-529-42
: Sequence 42, Application US/09300529
: Patent No. 6066783
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalin M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6066783artls Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.308
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/300,529
: FILING DATE: TBA
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/469,334
: FILING DATE: 06-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/463,483
: FILING DATE: 05-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993

```

Query Match	0.9%	Score 43	DB 3	Length 1241
Best Local Similarity	44.7%	Pred. No. 0.2		
Matches 215	Conservative	0	Mismatches 260	Indels 6
				Gaps 1

QY	2972	AGAAAGCTGGGAAGGAGCAATCAATAAATACCTGAAGAAGACATCTGCAACTGAGCTTCGACCTGT	3031
Db	88	AGAAAGGAGAAAGAGTGGAAAGCTTACCGCCACCGAGAAAGGCAAGATGAAACAACCTTCTCTGG	147
QY	3032	ACACAAGATGCACAATCTGAAACTGGAGAGACTTTCGACAAGAAGAACAATCGGGCAAGT	3091
Db	148	ACACAAGAAACGATCATCAAGACCACCACTACAGAGATCACTTTCAGCATATAGCCGGCAGCT	207
QY	3092	ATTAAGATGCGATGCATGAAGATTGACTCTCTGCTCAAGAGACAGTTGA-----AAGCAAC	3145
Db	208	TCGAGGACGAGATCAACAGACTTGAAGAGATGACACAAGTGTTCGACAAGACCAACTGTA	267
QY	3146	TCAACTCACTGAAACAATTCGGAAAAACAGTACTGACAGAACTTCTCACTGTGTTCCTCAACAAGA	3205
Db	268	GCAACAGCATCATCACTCACTCAAGAAAGTGGAGCCCAACCACTCGGCTTCAACAAGAGCC	327
QY	3206	AGAAAGAAAGCGGAGATCGCGGAGACAGAGAAACAATCTTGGAGAACAACAAGATTCTTCA	3265
Db	328	TGACCGAGGGCAACACCATCAACAGGAGCGCATGCGCCAGTTTCAAGAGAGCAAGTTCCTG	387
QY	3266	AATCACTACAAGAGCCCTCGTCAAGTATTATTAATGGGAGTCTTCTCTGGAAGACTCTCT	3325
Db	388	ACCGCGCATCAAGTTTCGACAGCTTACTCTGGACACCCACTGACCGCCACGAGGTGAGACA	447
QY	3326	CCGAGAGAGCATTCACAGACCGAGATTAATACGCCACCTCGAACACTTCAAGTCTCTGT	3385
Db	448	GCAAGAGAGCGCGTGATCTCGAAGGTATCCAGCTCCCAAGCGGCAAGGGCAGACACACCCCA	507
QY	3386	CTTAAGCTCGAAGGCAAGCTGAAGAGAACCTGAGACTGGAGAGAAGAAGCTCACTACC	3445
Db	508	CCAAAGCGGGCGTGATCTCTGAAACAACAGCAAGTACAAAGATGCTGATTCGACAACAAGCTTCA	567
QY	3446	T 3446	
Db	568	T 568	

RESULT 54
US-08-471-033-45
; Sequence 45, Application US/08471033
; Patent No. 5770656
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozietel, Michael G

```

APPLICANT : Mullins, Martha A
APPLICANT : Nye, Gordon J
APPLICANT : Carr, Brian
APPLICANT : Desai, Nalini M
APPLICANT : Kostichka, N. Kristy
APPLICANT : Duck, Nicholas B
APPLICANT : Estuch, Juan J
TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SRO ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 1358 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
FEATURES:
NAME/KEY: CDS
LOCATION: 9..1355
OTHER INFORMATION: /note= "Maize optimized VIF2A(a)
with the Bacillus secretion signal removed and the vacuola
OTHER INFORMATION: targetting signal inserted as contained in pCIB5533"
US-08-471-033-45

Query Match          0.9%: Score 43; DB 1; Length 1358;
Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 2972 AGAGCTGTGAAGAGACATCATTAAGCTGAAAGACACTGCATCACTGAGCTTCGACCTGT 3031
      ||||| | ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 205 AGAAGAAGAAAGATGTGAGCTTACC GCCACCGAAGAGGGCAABATGATAACAACACTTCC TCG 264

QY 3032 ACAACAAGTACCAAACCTGAACCTGGAGAGACTCTTGACACAAGACAAGACACTCGGCAAGT 3091
      ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 265 ACACACAGAAGACGACATCAAGGCCAACATAACAAGAGATCACCTTCAGATATACCGGAGCT 324

QY 3092 ATTAAGATGCAGATCAAGAAGTTGACTGCTGCTCAAGAGAGACGACTTGA-----AAGCAAC 3145
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 325 TCGAGGACGAGATCAAGGACCTGAAAGAGATCGACAAGATGTTTGCACACAAGACCAACTGA 384

```



```

APPLICANT: Kostichka, N. Kristly
APPLICANT: Duck, Nicholas B
APPLICANT: Estreich, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spittell, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 1358 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9...1355
OTHER INFORMATION:
OTHER INFORMATION: /note= "Maize optimized VIP2A(a)
with the Bacillus secretion signal removed and the vacuolar
targeting signal inserted as contained in pcIB5533"
US-08-463-483A-45

```

	Query Match	0.9% Best Local Similarity	Score 43	DB 2	Length 1358
	Matches	215	Conservative	0	Mismatches 260; Indels 6; Gaps 1;
QY	2972	AGAGACTGGAGAGGACATCATTAAGCTGGAAGACACTGCACACTGAGCTTCGACCTGT	3031		
Db	205	AGAAAGGAGAAAGAGCTGAGAGCTTACCCGACCAGAAAGGCAAGATGAAACAACCTTCGG	264		
QY	3032	ACAACAAGTACAAACTGAAACTGGGAGACACTCTTGACAGAAAGAAAGACAGCGGCAGT	3091		
Db	265	ACAAACAAAGACACTCAAGACCACTCAAGAGAGATACCTTCAGCATAGCGGCGAGCT	324		
QY	3092	ATAAGATCGAGATCAAGAAGTTGACTCTGCTCAAGAGACAGCTTGA-----AAGCAAAAC	3145		
Db	325	TCGAGGAGGAGATCAAGAGCTGGAAGGAGATGACAAAGATGTTGCAACAAGACCAACTGA	384		
QY	3146	TCAACTCACTGAACAATCCGAAACACGTAATCTGAGAACTTCTCAAGTGTCTTCAACAAAGA	3205		
Db	385	GCAACAGCATATATCCCTTACAAAGACGTAAGAGCCACCAACCATTCGGCTTCAACAAGAGCC	444		

[illegible]

RESULT 57
 US-08-471-046A-45
 Sequence 45, Application US/08471046A.
 Patent No. 5866326
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Kozien, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalini M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
 TITLE OF INVENTION: Protein Genes
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5866326artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,046A
 FILING DATE: 05-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SQLW4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 1358 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1355 /note= "Maize optimized VIP2A(a)
OTHER INFORMATION: with the Bacillus secretion signal removed and the vacuolar
OTHER INFORMATION: targeting signal inserted as contained in pCIB5533"
US-08-471-046A-45

Query Match 0.98; Score 43; DB 2; Length 1358;
Best Local Similarity 44.7%; Pred. No. 0.21;
Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 2972 AGAAGCTGGAAGAGGAGCATATTAAGCTGGAAGAGACACTGCAAGTGGCTTGCACCTGT 3031
DB 205 AGAAGGAGAGAGGAGTGAAGCTTACCGCCACCGAGAGGCGCAAGATGAGCAACTTCTCTG 264
QY 3032 ACAACAAGTACAACAGTGAAGTGAAGAGTCTTTCGACAAAGAGAGACAGTGGCAAGT 3091
DB 265 ACAACAAGAGAGCATCAAGACCACTACAAAGAGATCACTTCAGATATGCGGGAGCT 324
QY 3092 ATTAAGATGAGATCAAGAACTTGAATCTGCTCAAGAGACAGCTTGA-----AAGCAAC 3145
DB 325 TCGAGAGCGAGATCAAGAGCTGAAGAGATGCAAGAGATGTCGACCAAGACCAACTGA 384
QY 3146 TCAATCACTGAACAATCCGAAACAGTACTGAGAACTTCTAGTGTCTTCAACAGA 3205
DB 385 GCAACAGCATCATCACCCTACAAAGAGCTGAGGCCACCAACATCGGTTCAACAAGAGCC 444
QY 3206 AGAAGGAGAGCGGAGATGCGCGGAGACAGAGACACTCTGAGAAACCAAGATTCTTCTCA 3265
DB 445 TGACCGAGGGCAACACCATCAACAGCGCCATGCGCCAGTTCAAGAGACAGTCTCTCT 504
QY 3266 AACACTACAAAGGCTGCTCAATATTATATGCGCAGTCTTCTCTCTCAACACTCTCT 3325
DB 505 ACCGACACATCAAGTTCGAGAGTACCTGAGACCCACCTGACCGCCACAGAGTGAGCA 564
QY 3326 CCGAGGAGAGCATCAAGACGAGATTAACCTGACCGACCTCGAAGACTTCAAGGCTCTGT 3385
DB 565 GCAAGGAGCGCGTGAATCTTAAGGTGACCTCCCGAGCGCAAGGCGACGACCAACCCCA 624
QY 3386 CTAAAGCTGGAAGGCAAGCTGAAGAGCAACCTGAACTGGAAGAGAAAGAACTGCACTACC 3445
DB 625 CCAAGCGCGCGGTGATCTTAACACAGAGATTAACAAGATGCTGATGCAACAGCGGTACA 684
QY 3446 T 3446
DB 685 T 685

RESULT 58
US-08-470-566B-45
Sequence 45, Application US/08470566B
Patent No. 5872212
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziele, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J

TITLE OF INVENTION: No. 5872212e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 1358 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1355 /note= "Maize optimized VIP2A(a)
OTHER INFORMATION: with the Bacillus secretion signal removed and the vacuolar
OTHER INFORMATION: targeting signal inserted as contained in pCIB5533"
US-08-470-566B-45

Query Match 0.98; Score 43; DB 2; Length 1358;
Best Local Similarity 44.7%; Pred. No. 0.21;
Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 2972 AGAAGCTGGAAGAGGAGCATATTAAGCTGGAAGAGACACTGCAAGTGGCTTGCACCTGT 3031
DB 205 AGAAGGAGAGAGGAGTGAAGCTTACCGCCACCGAGAGGCGCAAGATGAGCAACTTCTCTG 264
QY 3032 ACAACAAGTACAACAGTGAAGTGAAGAGTCTTTCGACAAAGAGAGACAGTGGCAAGT 3091
DB 265 ACAACAAGAGAGCATCAAGACCACTACAAAGAGATCACTTCAGATATGCGGGAGCT 324
QY 3092 ATTAAGATGAGATCAAGAACTTGAATCTGCTCAAGAGACAGCTTGA-----AAGCAAC 3145
DB 325 TCGAGAGCGAGATCAAGAGCTGAAGAGATGCAAGAGATGTCGACCAAGACCAACTGA 384
QY 3146 TCAATCACTGAACAATCCGAAACAGTACTGAGAACTTCTAGTGTCTTCAACAGA 3205
DB 385 GCAACAGCATCATCACCCTACAAAGAGCTGAGGCCACCAACATCGGTTCAACAAGAGCC 444

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 23:49:26 ; Search time 4855 Seconds

(without alignments)
16479.033 Million cell updates/sec

Title: US-09-269-874A-2

Perfect score: 4940

Sequence: 1 cgcacgcgtatgaataatcat.....ttcatcctaataatcgtatgg 4940

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
1	235	4.8	537	13	B1815756	B1815756 PFESTRoaa3
2	220.8	4.5	600	14	B0452315	B0452315 PFESTRoaa9
3	194	3.9	500	9	A0088128	A0088128 PFESTRoaa3
4	192.4	3.9	491	14	B0597179	B0597179 PFESTRoab3
5	179.8	3.6	432	14	B0633144	B0633144 PFESTRoab3
6	168.8	3.4	500	9	A0086832	A0086832 PFESTRoab3

C	7	161.2	3.3	501	13	B1814631	B1814631 PFESTRoaa4
	8	160.4	3.2	422	14	B0451709	B0451709 PFESTRoab0
	9	160	3.2	522	14	B0452445	B0452445 PFESTRoaa9
	10	159	3.2	455	14	B0451036	B0451036 PFESTRoab0
	11	158	3.2	500	9	A0087573	A0087573 PFESTRoaa3
	12	156.4	3.2	500	9	A0088129	A0088129 PFESTRoaa3
	13	134.4	2.7	303	13	B1815239	B1815239 PFESTRoaa1
	14	132.4	2.7	419	13	B0451367	B0451367 PFESTRoaa9
	15	124.4	2.5	408	13	B1814931	B1814931 PFESTRoaa0
	16	124.4	2.5	418	14	B0451572	B0451572 PFESTRoab0
	17	124.4	2.5	422	14	B0451661	B0451661 PFESTRoab0
	18	119.8	2.4	365	14	N97742	N97742 1183C3 czap
	19	112.2	2.3	597	13	B1815959	B1815959 PFESTRoab0
	20	108.8	2.2	500	9	A0086246	A0086246 PFESTRoaa3
	21	108.8	2.2	755	13	B18159731	B18159731 PFESTRoaa4
	22	107.6	2.2	753	13	B18160423	B18160423 PFESTRoaa4
	23	105	2.1	678	13	B18165725	B18165725 PFESTRoaa4
	24	102.6	2.1	402	13	B1670649	B1670649 PFESTRoaa0
	25	101.6	2.1	400	14	N97689	N97689 1114C3 czap
	26	98.2	2.0	354	14	N97605	N97605 1039C3 czap
	27	97.2	2.0	647	17	A2526918	A2526918 261PbC05
	28	91.2	1.8	660	13	B18159472	B18159472 PFESTRoaa9
	29	91.2	1.8	777	13	B18170928	B18170928 PFESTRoaa9
	30	89.6	1.8	732	13	B18160106	B18160106 PFESTRoaa9
	31	88.4	1.8	313	14	T18122	T18122 0676C3 czap
	32	86.2	1.7	362	14	B0451205	B0451205 PFESTRoab1
	33	85.2	1.7	352	14	B0577302	B0577302 PFESTRoab1
	34	82.8	1.7	282	14	T18003	T18003 0325C3 CBSP
	35	80.6	1.6	878	17	B18153470	B18153470 PFESTRoaa9
	36	79	1.6	483	13	B18165179	B18165179 PFESTRoaa9
	37	79	1.6	703	13	B18167609	B18167609 PFESTRoaa9
	38	79	1.6	769	13	B18166688	B18166688 PFESTRoaa9
	39	78.8	1.6	1147	13	B13042	B13042 T30M24-SP6
	40	77.2	1.6	503	13	B18156118	B18156118 PFESTRoaa9
	41	74.4	1.5	344	13	B18159366	B18159366 PFESTRoaa9
	42	73.8	1.5	540	17	A2813205	A2813205 2M0080C24
	43	73.4	1.5	948	13	B1815494	B1815494 OP20570 M
	44	71.6	1.4	1885	10	BE420745	BE420745 HMM002-B0
	45	69.8	1.4	691	17	BH040136	BH040136 RRC1-24-2
	46	68.6	1.4	682	13	B18170771	B18170771 PFESTRoaa9
	47	68.6	1.4	908	17	A2548467	A2548467 PFESTRoaa9
	48	68.4	1.4	544	13	B181420171	B181420171 PFESTRoaa9
	49	68.4	1.4	596	17	A2460541	A2460541 1M0265124
	50	68	1.4	907	17	CNS02134	CNS02134 PFESTRoaa9
	51	67.2	1.4	494	13	B18160881	B18160881 PFESTRoaa9
	52	67.2	1.4	500	9	A0087665	A0087665 PFESTRoaa9
	53	66.4	1.3	827	17	CNS02156	CNS02156 PFESTRoaa9
	54	66	1.3	1168	14	B0931326	B0931326 PFESTRoaa9
	55	65.8	1.3	731	17	A2191902	A2191902 SP.1020_A
	56	65.6	1.3	960	17	CNS005F3	CNS005F3 PFESTRoaa9
	57	65.4	1.3	626	13	B18126817	B18126817 PFESTRoaa9
	58	65.4	1.3	783	17	CNS01112	CNS01112 PFESTRoaa9
	59	65.4	1.3	897	17	CNS07AB2	CNS07AB2 PFESTRoaa9
	60	65.4	1.3	1135	17	CNS033C0	CNS033C0 PFESTRoaa9
	61	65.2	1.3	1055	14	B0876453	B0876453 PFESTRoaa9
	62	65	1.3	885	17	A2541857	A2541857 PFESTRoaa9
	63	64.8	1.3	641	17	A0946120	A0946120 PFESTRoaa9
	64	64.8	1.3	1101	17	CNS00KB5	CNS00KB5 PFESTRoaa9
	65	64.4	1.3	500	9	A0086975	A0086975 PFESTRoaa9
	66	64.4	1.3	843	17	A2551618	A2551618 PFESTRoaa9
	67	64.2	1.3	1223	17	B12981	B12981 T24D11-SP6
	68	64	1.3	781	17	AG171177	AG171177 PFESTRoaa9
	69	64	1.3	1309	13	B18163041	B18163041 PFESTRoaa9
	70	63.8	1.3	976	17	AG16173	AG16173 PFESTRoaa9
	71	63.6	1.3	650	10	AV995927	AV995927 PFESTRoaa9
	72	63.6	1.3	1084	17	AG0322755	AG0322755 PFESTRoaa9
	73	63.4	1.3	817	13	B18163917	B18163917 PFESTRoaa9
	74	63.2	1.3	585	17	A2542280	A2542280 PFESTRoaa9
	75	63	1.3	356	13	B18137970	B18137970 PFESTRoaa9
	76	63	1.3	910	13	B1815636	B1815636 PFESTRoaa9
	77	63	1.3	932	17	A2689160	A2689160 PFESTRoaa9
	78	62.8	1.3	443	6	BM374465	BM374465 PFESTRoaa9
	79	62.8	1.3	1300	13	BM468018	BM468018 PFESTRoaa9

80	62.6	1.3	603	13	BJ366946
81	62.6	1.3	997	17	CNS0134P
82	62.6	1.3	1036	17	CNS0134P
83	62.4	1.3	394	9	AU060224
84	62.4	1.3	439	13	BM160668
85	62.4	1.3	527	13	BM160302
86	62.4	1.3	780	13	BM166681
87	62.4	1.3	754	13	BM166681
88	62.2	1.3	1101	17	CNS0039G
89	62.2	1.3	804	17	AG077527
90	61.8	1.3	848	17	AZ672705
91	61.8	1.3	870	17	AZ672705
92	61.6	1.2	671	17	AG030286
93	61.6	1.2	681	17	CNS02EOD
94	61.4	1.2	538	13	BJ417399
95	61.2	1.2	617	13	BJ416228
96	61.2	1.2	839	17	CNS0122R
97	61.2	1.2	359	13	BJ366856
98	61.2	1.2	867	17	CNS0054A
99	61.2	1.2	871	17	AZ671726
100	60.8	1.2	576	9	AU060772
101	60.8	1.2	634	17	AZ008814
102	60.8	1.2	783	17	CNS011P
103	60.8	1.2	1101	17	CNS0153V
104	60.6	1.2	744	17	AG044338
105	60.6	1.2	877	17	AZ531291
106	60.6	1.2	912	17	AZ551092
107	60.4	1.2	648	10	AV676742
108	60.4	1.2	913	17	BM158934
109	60.4	1.2	997	17	CNS005TE
110	60.2	1.2	461	17	AG072942
111	60	1.2	950	13	BJ371297
112	60	1.2	834	17	BI2387
113	60	1.2	964	13	BM416130
114	60	1.2	966	13	BM415686
115	60	1.2	1184	17	BI3117
116	60	1.2	1398	13	BM416038
117	59.8	1.2	627	10	BJ426316
118	59.8	1.2	688	10	AV671840
119	59.8	1.2	884	17	AZ199419
120	59.8	1.2	884	17	AG056518
121	59.8	1.2	896	13	BM415641
122	59.8	1.2	924	13	BM415947
123	59.6	1.2	1008	17	AG134526
124	59.4	1.2	538	17	AZ064820
125	59.4	1.2	582	13	BJ369186
126	59.4	1.2	697	17	AZ374916
127	59.4	1.2	787	17	CNS007XT
128	59.4	1.2	959	17	CNS00655
129	59.4	1.2	1491	17	BE882936
130	59.2	1.2	877	17	AZ548629
131	59	1.2	858	17	CNS0127J
132	59	1.2	1185	12	BF273407
133	58.8	1.2	814	17	AZ672545
134	58.8	1.2	947	17	AG056380
135	58.8	1.2	954	14	BO670925
136	58.6	1.2	613	13	BJ360997
137	58.6	1.2	640	13	AZ476780
138	58.6	1.2	784	9	AL5525973
139	58.6	1.2	931	17	AG446286
140	58.6	1.2	952	17	CNS014BF
141	58.4	1.2	803	17	BI1974
142	58.4	1.2	865	17	AQ324474
143	58.4	1.2	1007	17	CNS006X5
144	58.2	1.2	421	17	AZ113646
145	58.2	1.2	850	17	AZ186328
146	58	1.2	546	14	C22974
147	58	1.2	784	14	BO962626
148	58	1.2	861	17	AZ137643
149	58	1.2	866	17	AG073614
150	58	1.2	934	13	BM415088
151	58	1.2	979	13	BM415502
152	58	1.2	1204	13	BM452144

153	57.8	1.2	539	9	AU076365
154	57.8	1.2	542	13	BM167382
155	57.8	1.2	816	17	CNS02KJV
156	57.8	1.2	878	17	CNS0028X
157	57.8	1.2	886	17	CNS075DF
158	57.8	1.2	931	17	BM160272
159	57.8	1.2	952	17	AO897460
160	57.8	1.2	1042	17	CNS0148K
161	57.8	1.2	1059	17	CNS0022B
162	57.8	1.2	1079	17	CNS0351X
163	57.8	1.2	1101	17	CNS001JT
164	57.6	1.2	297	13	BM633446
165	57.6	1.2	608	13	BJ330059
166	57.6	1.2	617	17	AZ296172
167	57.6	1.2	735	17	AZ185710
168	57.6	1.2	888	17	CNS027EX
169	57.6	1.2	964	17	AZ667479
170	57.6	1.2	1215	14	BM822560
171	57.4	1.2	739	14	BO991717
172	57.4	1.2	829	17	CNS011NU
173	57.2	1.2	571	17	AZ952930
174	57.2	1.2	792	17	AG032727
175	57.2	1.2	796	17	CNS0118D
176	57.2	1.2	944	17	CNS0025G
177	57.2	1.2	1139	17	AO897537
178	57.2	1.2	1382	14	BM609864
179	57	1.2	645	9	AI389106
180	57	1.2	870	17	AO866797

ALIGNMENTS

BI815756	537 bp	mRNA	linear	EST 03-OCT-2001
PF850aa32c09.y1	Plasmodium falciparum 3D7 asexual cDNA	Plasmodium falciparum		
fatcplparum cdna 5'	similar to TR:Q9YV68 Q9YV68 MSAL PROTEIN			
sequence.				
BI815756	BI815756.1	GI:15909334		
malaria parasite P. falciparum.				
Plasmodium falciparum				
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
1 (bases 1 to 537)				
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,				
Marta, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,				
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,				
Tsagaris, V., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,				
Waterston, R., Wilson, R. and Sibley, D.				
Washu Plasmodium EST Project				
Unpublished (2001)				
Contact: L. David Sibley				
Washu Plasmodium EST Project				
Washington University School of Medicine				
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA				
Tel: 314 286 1800				
Fax: 314 286 1810				
Email: estewatson.wustl.edu				
Library was constructed by Deoban Chakrabarti DNA sequencing by:				
Washington University Genome Sequencing Center For information on				
obtaining a clone please contact: L. David Sibley				
(sibley@bcm.wustl.edu), Washington University				
Seq primer: -40UP from Gibco				
High quality sequence stop: 422.				

FEATURES

source

1. 537
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
 /lab_host="DH10B (genehog, Invitrogen, Inc.)"

/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 Zapit vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXassist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 217 a 75 c 83 g 162 t
 ORIGIN

Query Match 4.8%; Score 235; DB 13; Length 537;
 Best Local Similarity 66.0%; Pred. No. 9.3e-43;
 Matches 356; Conservative 0; Mismatches 180; Indels 3; Gaps 1;

QY 4401 AACATCCAGACAGCTGCGAGATTTCAGAAAAATACAAATTCGTCGCAATTGCAGA 4460
 Db 1 AGCAATTGATGACAAAATAGATCTTTTAAAAACCTTACGACTCGAAGCAATTTAAAA 60
 QY 4461 CCTGCTACCATTTTAACACACACATCTCTGACCAAGTTTCTGCTCACTGGCATGT 4520
 Db 61 ATTGTAATATATACGAAAAAAGATATGCTGGCAAAATTAATTAAGTACGAGTATGT 120
 QY 4521 GTTCGAAAACCTCGCAAAAGCTGCTGAGCAATCTGCTGCGGCAACCTGCGAGGAT 4580
 Db 121 TC-----AAAATTTTCTTAATACATAATATTCAAAATTAATTGAGAAAATTCAGATAT 177
 QY 4581 GCTGAACATCTCCAGACCAATGCGTGAAGAAACAGTCCCGCAATAGCGCTGT 4640
 Db 178 GTTAAACATTTACACACCAATGCGTGAAGAAACAGTCCAGAAAATTCGATGTTT 237
 QY 4641 CAGCATCTGACAGCGCCGGAAGAGTGTCTCTGAACATTCACAAAGAAAGAGA 4700
 Db 238 CAGACATTTAATGAAAGAAAGATGTAAATGTTTAAATTAATTAACAAAGAGGTGA 297
 QY 4701 TAAATGCTGGAGAACCAACCTGACTGCAATGAAGAAACATGCGGCTGAGCGCGA 4760
 Db 298 TAAATGCTTGAATAATCCAAATCTACTTGTACGAAAATAATGATGATGATGCGAGA 357
 QY 4761 TGTAAATGACGAGAGAGACAGCGCTCTAACGGAAGAAATACATGCGAGTGTAC 4820
 Db 358 TGCCACATGTACCGAAGAAAGATTGAGTAGCAGCAAGAAATACATGATGATGTAC 417
 QY 4821 TAAACCCGACCTCTATCCACCTTCGACGAGATTTTTCGTCACCTTAATTTCTCTGG 4880
 Db 418 TAAACCTGATTTCTTAATCTTTCGATGATGATTTTTCGACCTTCTCTCTTGG 477
 QY 4881 CATCTCTCTGCTGATCTCATGCTGATCTGATGATCTGATGATGATGATGATGATG 4939
 Db 478 AATATCATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 536

RESULT 2
 BO452315
 LOCUS 600 bp mRNA Linear EST 29-MAY-2002
 DEFINITION Pfesr0ae95a12.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
 falciparum cDNA 5' similar to SW:MSPI_PLAF3 PI9598 MERZOITE
 SURFACE PROTEIN 1 PRECURSOR ;, mRNA sequence.
 ACCESSION BO452315
 VERSION BO452315.1 GI:21255427
 KEYWORDS EST.
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 REFERENCE 1 (bases 1 to 600)

AUTHORS Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,
 Maria, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Thieling, B.,
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,
 Tsagarelis, V., R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,
 Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,
 Waterston, R., Wilson, R., and Sibley, D.
 TITLE Washu Plasmodium Est Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: L. David Sibley
 Washu Plasmodium Est Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Library was constructed by Debopam Chakrabarti DNA sequencing by:
 Washington University Genome Sequencing Center for information on
 obtaining a clone please contact: L. David Sibley
 (sibley@borcim.wustl.edu), Washington University
 Seg primer: -40UP from Gibco
 High quality sequence stop: 432.
 Location/Qualifiers
 1..600
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /clone.lib="Plasmodium falciparum 3D7 asexual cDNA"
 /lab_host="DH10B (Genetec, Invitrogen, Inc.)"
 /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 Zapit vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXassist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 276 a 79 c 72 g 173 t
 ORIGIN

Query Match 4.5%; Score 220.8; DB 14; Length 600;
 Best Local Similarity 63.6%; Pred. No. 1.7e-39;
 Matches 336; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 3064 TTGCAAGAAAGACAGCTGCGCAAGTAAAGATGACAGTAAAGATGATGCTGCTC 3123
 Db 1 TTTAATAAGAAAAAGAACTGCGCAAGCAAAATTAATAAACTTAATTAATA 60
 QY 3124 AAGAGACGCTTGAAGAGAACTCACTGACGAACTCGAAGACACTGAGAAC 3183
 Db 61 AAAGACATTAAGATCAAAATTAATGATTAATTAATGACGATTAATTAATTAATTA 120
 QY 3184 TTCTCAGTCTTTCACAG 3243
 Db 121 TTTTCTGTTTCTTAAACAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 QY 3244 GAGAACACCAAGATTTCTGCAACACATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 3303
 Db 181 GAAACACAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240
 QY 3304 TCTTCTCCTGAG 3363
 Db 241 TCATCTCCATTAATAAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 QY 3364 CTCGAGAACTTCAAGAGTCTGTAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3423
 Db 301 TTAGAAAAATTTAGAGTATTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTA 360

QY 3424 GAGGAGGAAGACCTCAGGTACTCTCTCTGCGGACATGCATCACTCATGCGCCAGTCCAAG 3483
| | | | | | | | | | | | | | | | | |
Dd 361 GGAAAGAAAAATTTCCTTTCTTATCAAGTGCAATTACATCATTTAATTACTGAAATTAATA 420
| | | | | | | | | | | | | | | | | |
QY 3484 GAAGTCATTTAAGAACAAGAACATCACCGGCATATGCCCAAGGAGAATAATATACAGACGTG 3543
| | | | | | | | | | | | | | | | | |
Dd 421 GAAGTAAATTAATAAAAATTAATAAATTTATACAGTAATTTCTCCAAAGGAAAAATATATAGAAAGTT 480
| | | | | | | | | | | | | | | | | |
QY 3544 AATAACGCACTGGATCTTACAGAAGAGTTCCCTGCCCTGAAGAACAGAT 3591
| | | | | | | | | | | | | | | | | |
Dd 481 AACGAAGCTTTAAATCTTACCAAATTTTTCTCCCCAAGAACAAAGTT 528

RESULT 3
AUT088128

LOCUS	500 bp	mRNA	linear	EST	27-JAN-2001
AU088128					
DEFINITION	Sugano	Malaria	CDNA library	Plasmodium falciparum	307 CDNA

for precursor of major merozoite surface antigens, mRNA sequence.
AU088128

KEYWORDS EST.
SOURCE *Plasmodium falciparum* 3D7.

REFERENCE
1 (bases 1 to 500)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

TITLE
FULL-malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum

MEDLINE 20574754
COMMENT Contact: Junichi Watanabe

Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

TEL: 01-3-5449-5570
Fax: 81-3-5449-5410
Email: twatanabemanager.lms.u-tokyo.ac.jp

SUZUKI, I., IOSHIMIZU-MAKIGAWA, N., MAIYAMA, A., SUYAMA, A. and SUGANUMA, S. Construction and characterization of a full length-enriched and/or a 5'-end-enriched cDNA library Gene 200 (1-2): 149-156 (1997).

SOURCE	FEATURES
1. .500	location/qualifiers
/organism="Plasmodium falciparum 307"	

```

/strain="JD/"
/db_xref="taxon:36329"
/oligon="vnp-6550"

```

```
clone_lib="Sugano Malaria cDNA library"
dev_stage="erythrocytic stage"
```

ORIGIN

Best Local Similarity	64.48;	Pred. No. 2.1e-33;
Matches 290; Conservative	0;	Mismatches 160; Indels 0; Gaps 0;

3142 A A A C T C A A C T G A A C A A T C C G A A A C A C G T A C T G C A G A A C T T C T C A G T G T T C T T C A A C 3201

DY 3302 AAGAGAAGGAAACCGATCCGCAGACAGAGAACACTCTGGAGAACACCAGAATTCCTT 3361

DU C AAAAAGAAAACAATTAACCCCAATAATGTATATTTCAATTATCIIICIIIAAC 04

Db 65 AAAAAAAAAAGAGCTGAAATAGCAGAACTGAAACACATTTAGAAAAACAAAAATATTTA 124

[illegible]

3322 CTTCCGAGGAGACATCCAGACGAGATACTACGCCAGCCTCGAGAACTTCAAGTTC 3381

Db 185 TTAAGTGAAGTATCAATTCAAACGAAGATAATTTATGCCAATTTAGAAAATTTAGAGTA 244

Qy	3442	TACCTCTCTACGGCATGCACTACCTGATCGCCAGCTCAAGAAAGTCA	TTAAGAACAG	3501
Db	305	TTCTTATCAAGTGGATTTACATCATTTTAATTAAGTAATTTAAAGAAAGTAA	TTAAAAATTA	364
Qy	3502	AACATACACCGGCAATAGCCCCAGCAAGAAATATACAGACGTGAATATACGCACTGGAATCT		3561
Db	365	AATTATACAGGTATTTCTCCAAAGTGAATAATATAGAAAGTTATACGAAAGCTTTAAATCT		424
Qy	3562	TACAGAAGTTCCTGCTCGATGAAGAACAGAT		3591
Db	425	TACGAATAATTTTCTCCAGAAAGCAAAAGTT		454

RESULT 4
DOE07170

LOCUS	491 bp	mrna	linear	EST 24-JUN
BQ597179				
DEFINITION	PfSTEoab35h07.y1 Plasmodium falciparum 3D7 asexual		CDNA	Plasmoc

SURFACE PROTEIN 1 PRECURSOR ;, mRNA sequence.
 BQ597179

KEYWORDS EST.
SOURCE malaria parasite *P. falciparum*.

REFERENCE
1 (bases 1 to 491)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

Marras, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, E.
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko,

Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.

unpublished (2001)
Contact: L. David Sibley
Journal of Management Inquiry 11(1)
DOI: 10.1177/1056492602237502

Washnu Prastimodum Esi Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
Fax: 314 286 1810
Email: osr@watson.wustl.edu

Library was constructed by Depopam Chakraborti DNA sequencing Center Washington University Genome Sequencing Center For information obtaining a clone please contact: Dr. David Sibley

(sibley@borcim.wustl.edu), Washington University
Seq primer: -400p' from Gibco

FEATURES	Location/Qualifiers
source	1. .491

```

/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"

```

/note="Vector: pBluescript SK plus; Site_1: EcoRI; site_2: XhoI; library was constructed by Debopam Chakrabarti.

total RNA samples were isolated from mixed stage saproin(0.1%) - lysed *P. falciparum* 3D7 infected erythrocytes by the acidic guanidium-phenol chloroform

mRNA isolation system (Promega, WI) using streptavidin magnetic beads. Directional cDNA libraries were constructed using the

constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the

cdna inserts in the library was between 1.0 and 1.5 kb. Clones were mass excised using the Exassist helper phage

8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

Matches : 288; Conservative : 0; Mismatches : 195; Indels : 3; Gaps : 1;

```
OY 1533 CTTGAGAGAAATGATGGAGAAAGCGTACTCTTACAAAGTGGGAGAACTGACACACCATAA 1592
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 CGTAGATAAATATTCATTCGACAGATATACATATATGTTGAAAACAAAGTATATATTA 74
OY 1593 TACCTTTCATCTCATGAGAAATTCAGCATATCTTGAAGAGCTCACCAGGCTTTAA 1652
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 TAAATTTTCATCTCATATATTCGTATATATGTTCAAAAATTTAAAAAGGCTCTTTC 134
OY 1653 GTATATGAGAGACTATCTCTGCGGAGACATTTGTTGGAGAAAGCACTAAAGTATTACA 1712
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 ATATCTTGAAGATTTATCTTTTAAAGAAAGAAATTTCTGAAAAATTTTAAATCATATTA 194
OY 1713 GAATCTCATAGTAAAGTGAAGAAAGAGATGAGACGCTGTTTGAAGACATTAAGAGAA 1772
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 TACTTTGAAAACCTGGCCCTCGAAGCTGATATATAAAAATTTACAGAGAAATTAAGAGTNG 254
OY 1773 TGAAGACAGTTGTTTGAAGAAAGATTTACAAAAGCAGAAATTAACGAGATGAGAGAT 1832
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 TGAAGACAAAATTCAGAAAAAAATTTTAAAGACTAACACATTCAGCAAAATG--GTTG 311
OY 1893 CCTGAGAGTCTCCGATATCTGTTAAAGTCCAGAGTGCAGAGAGTGCCTCATGACAAAGT 1892
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 CTAGCAAGTATCTGATATTTGTTAAATTTACAGTACAAAAGTTTATTAATTTAAAAAT 371
OY 1893 TGATGAACTCAAGAGACTCAACTCATCTGAGAGAGCTGAGTTAAACATATATATACA 1952
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 AGAAGACTTTAAGAAAGATAGAAATTTTAAAGAAATGACACACTAAAGATAGATTTCA 431
OY 1953 TGTGCGCAATAGTTATTAAGCAGAGAGATTAAGCAGAACCATATACCTCATGCTATCA 2012
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 TGTACCAATATTTATTAACCAAAATTAACAGAGACCATATTTATTAATTTATTA 491
OY 2013 GAAAGA 2018
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 AAAAGA 497
```

RESULT 12
LOCUS AU088129 500 bp mRNA linear EST 27-JAN-2001
DEFINITION AU088129 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA
clone Xpfn6560 similar to P.falciparum (NF7) gene for mezozoite
surface antigen 1, mRNA sequence.
ACCESSION AU088129
VERSION AU088129.1 GI:12390270
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7.
ORGANISM Plasmodium falciparum 3D7.
REFERENCE Eukaryota: Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 500)
AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE FULL-malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum
Nucleic Acids Res. 29 (1), 70-71 (2001)
JOURNAL MEDLINE 20574754
COMMENT Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanabemange.jms.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
,S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
source
1..500
/organism="Plasmodium falciparum 3D7"
/strain="3D7"
/db_xref="taxon:36329"
/clone="Xpfn6560"

/clone_11b="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 197 a 65 c 58 g 144 t 36 others
ORIGIN

Query Match 3.2%; Score 156.4; DB 9; Length 500;
Best Local Similarity 57.6%; Pred. No. 8.1e-25;
Matches 250; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

```
OY 3161 ATCCGAACACCTACTGCGAAGAACTTCTCAGTCTCTTCAACAAGAAAGAGCGGAGA 3220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 AACCCCATATGTTATTTACAAAACCTTTCTGTTTCTTTAAACANTNAAAAAGAGTGAAA 116
OY 3221 TCGCGAGACAGAGAAACACTGCGAAGAACCAAGATTCCTTCTCAACACTCAAGAGCC 3280
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 TAGCAGAAACGTANANNANNCATTAAAAAACCAAGANTATATTGAAACATTTTAAAGGAC 176
OY 3281 TCGTCAAGTATTATATATGCGAGTCTCTCTCTGAGACTCTCCGAGAGAGACTTC 3340
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 TTGTAAATATATATATATGATGTAATCATCTCCATTAAAACTTTAAGTGAAGTACATTC 236
OY 3341 AGACCGAGATATACATACGCGACCTCGAGAACTTCAAGTCTGTCTAAGCTCGAAGCA 3400
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 AAACGAAAGATATATATATGCAATNTAGANANNNTAGATATTAAGTNAAATATAGAA 296
OY 3401 AGCTGAAGGACAACTGGAACCTGAGAAAGAGAGCTGACCTCTACGCGACTGC 3460
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 AACCTCAATGATATTATTACATTTAGAAGAAAAATTTATCTTCTTATCAAGTGATTC 356
OY 3461 ATCACCCTATGCGCGAGCTCAAGAGATGTCATTTAAGACAGAACTACACCGGCAATPAGCC 3520
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 ATCATTTAATTACTGAAATTAAGAAAGTTATTAANNTAANTATATACAGGTNAATCTGC 416
OY 3521 CAAGCGAATATATACACAGCTGATTAACGCACTGGAATCTTACAAAGAGTCCGCTCG 3580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 CAAGTGAATATATATAGAAAGTTAAAGCACTTNAAACTTCAAAAAATTTTTCNCA 476
OY 3581 AAGGACAGATGTC 3594
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 NAAGCAAAAGTTNC 490
```

RESULT 13
LOCUS B1815239 303 bp mRNA linear EST 03-OCT-2001
DEFINITION B1815239 PfEST0a16b11.y1 plasmodium falciparum 3D7 asexual cDNA Plasmodium
falciparum cDNA 5' similar to TR:Q90A18 Q90A18 SURFACE PROTEIN-1 ;
mRNA sequence.
ACCESSION B1815239
VERSION B1815239.1 GI:15908372
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota: Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 303)
AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Matta,M., Hillier,L., Martin,J., Wylie,T., Dente,M., Theising,B.,
Bowers,Y., Gibbons,M., Riter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagarelisvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
TITLE Mashu Plasmodium Est Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
Washu Plasmodium Est Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: L. David Sibley

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
RESULT 15 B1814931	B1814931 pfESTca08b05.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9UAI8 Q9UAI8 SURFACE PROTEIN-1 ; mRNA sequence.	408 bp	mRNA	linear	EST 03-OCT-2001					
DEFINITION	B1814931 pfESTca08b05.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9UAI8 Q9UAI8 SURFACE PROTEIN-1 ; mRNA sequence.	408 bp	mRNA	linear	EST 03-OCT-2001					
ACCESSION	B1814931									
VERSION	B1814931.1	GI:15907780								
KEYWORDS	EST.									
ORGANISM	Plasmodium falciparum									
REFERENCE	1 (bases 1 to 408)									
AUTHORS	Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Peppe, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, R., Rilter, E., Bennett, J., Jenkes, E., Ronko, I., Tsagaris, S., Belagorod, L., Franklin, C., Carr, L., Grow, A., Mawer, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterson, R., Wilson, R. and Sibley, D.									
TITLE	Washu Plasmodium EST Project									
JOURNAL	Unpublished (2001)									
COMMENT	Contact: L. David Sibley Washu Plasmodium EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University putative full length read vector to vector length is 409 Seq primer: -40UP from Gibco.									
FEATURES	Location/Qualifiers									
Source	1. 408 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /clone_lib="Plasmodium falciparum 3D7 asexual cDNA" /lab_host="DH10B (GeneHog, Invitrogen, Inc.)" /note="vector: pBluescript SK plus; Site_1: EcoRI; Site_2 XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage zapping(0.1%) - lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin Magnesphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A+) RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXASist helper phage (Stratagene), the phagems were preclitpated with PEG 8000 and extracted with phenol/chloroform. Phagmid DNA was electroporated into DH10B cells."									
BASE COUNT	161 a 43 c 46 g 158 t									
ORIGIN										
Query Match	2.5%: Score 124.4; DB 13; Length 408;									
Best Local Similarity	72.5%: Pred. Mismatches 11;									
Matches 161;	Conservative 0; No matches 61; Indels 0; Gaps 0;									
4778	CAAAACCTACCTGACGATGAAACATGCGGGGTGACGCGCATGCTAAATGACCGAGG	4777								
1	CAAAATCTACTTGTAAACGAAATTAATGTTGGATGTGATGACATGCGCATATTTACCGAAG	60								
4778	AAGACAGCGGCTCTTACGGAAGAAATATCATATGCGAGTGTACTAAGCCGACTCTTAT	4837								
61	AAGATTACAGGTGACGACGAGAAATATCATATGGAATGTACTTAACCTGATTTCTTATC	120								
4838	CACCTCTTGACGGGATTTTTTGTCTCCAGCTTAATTTCCCTGGGACATCTTCCTTCGTCA	4897								

DB	121	CACITTTGCATGCTATTTTCTTCGACCTGCTCTTAACCTCTTACGATTCCTTATTA	180
QY	4898	TCCTCATGCTGATCCTGTACAGCTTCATCTAATAGATGCATG 4939	
DB	181	TACTCATGTTATATTATACACTTTCATTTAAAAATGTAGG 222	
RESULT	16		
LOCUS	B0451572	418 bp	mRNA
DEFINITION	PfSTEaob0b11.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;		
ACCESSION	B0451572		
VERSION	B0451572.1	GI:21254684	
KEYWORDS	EST.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Plasmodium falciparum		
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 418)		
TITLE	Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D., Marra, K., Hillier, L., Martin, J., Wyllie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagaris, N., Belagorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Ritchey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.		
JOURNAL	Unpublished (2001)		
COMMENT	Washu Plasmodium EST Project Contact: L. David Sibley Washu Plasmodium EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University Seq primer: -400p from Gibco. Location/Qualifiers 1. 418 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /clone_lib="Plasmodium falciparum 3D7 asexual cDNA" /lab_host="DH10B (GeneHog, Invitrogen, Inc.)" /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage erythrocytes by the acidic guanidium-phenol chloroform method. The poly A+ RNA was isolated by the polyA+-tract mRNA isolation system (Promega, WI) using streptavidin MagneSphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXASist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."		
BASE COUNT	171 a	43 c	46 g 158 t
ORIGIN			
Query Match	2.5%	Score 124.4;	DB 14; Length 418;
Best Local Similarity	72.5%;	Pred. No. 1.6e-17;	
Matches 161;	Conservative 0;	Mismatches 61;	Indels 0; Gaps 0;
QY	4718	CAAACCTACCTGCAGTAAGAAACATGGCGGTGTGAGCGCGATGCTAATGACCGGAGG 4777	
DB	1	CAAACTCCTACCTGTGAAGCAAAATTAATGTGTGATGATGACATGTCCACATGTACCGAAG 60	

Qy	4778	AAGCAGGGGCTCTACAGGAAGAAATACATGGCAGGTACTAGGCCGCTCCTATTC	4837
Db	61	AAGATTTCAGGTAGCGACGAGAAAGAAATACATGTGAATGTACTAATCGATTCTTATC	120
Qy	4838	CACCTTCCTGACGGGATTTTTTGGCTCCAGCTCTAAATTTCTCGGCAWCTCTCCGCTGA	4897
Db	121	CACTTTCGATGGATTTTCTGCTACGTTCTCTAATCTTTAGAAATATCATTTCTATTAA	180
Qy	4898	TCCATATGCTGATCTCTGATACGCTTCATCATATTAATGATG	4939
Db	181	TACTATGTATTATATTATACAGTTTCATTTAAAAATGTAG	222
RESULT 17			
LOCUS	BQ451661	422 bp	mRNA linear EST 29-MAY-2002
DEFINITION	PIES2ab07c09.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9DA18 Q9DA18 SURFACE PROTEIN-1 ;		
ACCESSION	BQ451661		
VERSION	BQ451661.1	GI:21254773	
KEYWORDS	EST.		
SOURCE	Plasmodium falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 422) Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Page,D., Marras,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsgareishvili,R., Belagorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Ritchey,J., Waddins,J., Kennedy,S., Levinso,D., Waterson,R., Wilson,R. and Sibley,D.		
TITLE	Washu Plasmodium EST Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: L. David Sibley Washu Plasmodium EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University Seq primer: -400P from Glbco. Location/Qualifiers		
FEATURES			
SOURCE	1..422 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /clone_lib="Plasmodium falciparum 3D7 asexual cDNA" /lab_host="DH10B (Genesorg, Invitrogen, Inc.)" /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyA+Treat mRNA isolation system (Promega, WI) using streptavidin MagneSphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."		
BASE COUNT	175 a	43 c	46 g 158 t
Query Match	2.5%	Score124.4;	DB 14; Length 422;

[illegible]

Db 43 ATGATGAATATCCCTCTCAGAGATTGAAATGATATGATATATATTTAAACCTTAA 102
 Oy 3877 GCAGCGCTATAGCTCTCTCAGAAACGCTGGAATAGCTGATCCCTCAATGTC 3936
 Db 103 GCTGAGATATATGAACTTTAAAAACAAATTTGAAAAACATTTCATTATTTAA 162
 Oy 3937 AACGTGAAGACATTCGTGAAACGCCCTTTAATAGAGAAAATTTCAAGAAGCTTG 3996
 Db 163 AATTGGAAGATATCTTAATTCAGCTCTTAAGAAAAGAAAATATTTCTTAATGATTA 222
 Oy 3997 GAGACGACCTTGATTCCTATTAAGACCTGACCTCCTCTACTACGTGTGCAAGACCA 4056
 Db 223 GAATCTGATTTAATGCAATTTAAACATATATCTCAATGATACATTTATGAAGATTA 282
 Oy 4057 TACAGTTCCTCAATGAAGAGAGAGGATTAATTTCTGTCTAGTACACTATATCAAG 4116
 Db 283 TTTAATATTTGAAATTCGAAACAAAACACACTTTTAAAAAGTTCAAAATATATAAAA 342
 Oy 4117 GACTCCATCGACACCGCATATCA 4139
 Db 343 GAATCAGTAGAAATGATATTTAA 365

RESULT 19
 Bm159597 597 bp mRNA linear EST 04-DEC-2001
 LOCUS Bm159597
 DEFINITION EST562120 PyBS Plasmodium yoelii yoelii cDNA clone pYCUG14 5' end,
 mRNA sequence.
 ACCESSION Bm159597
 VERSION Bm159597.1 GI:17305278
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Plasmodium yoelii yoelii.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 597)
 AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
 Fraser, C.M. and Carucci, D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlone@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES
 source
 1..597
 location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="pYCUG14"
 /clone_lib="PYBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: PAD-GAL4. At 20-25% parasitemia, blood was
 collected from BALB/cByd mice infected with Pyl17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HydrizAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was
 excised from the HydrizAP vector and plasmid DNA
 isolated."

Query Match 2.3%; Score 112.2; DB 13; Length 597;
 Best Local Similarity 52.4%; Pred. No. 1e-14;
 Matches 292; Conservative 0; Mismatches 238; Indels 27; Gaps 1;

BASE COUNT 268 a 76 c 92 g 161 t
 ORIGIN
 Oy 487 AACATTCATGCGTTCATAATATCTGATTGACGGTTACGAGAGATCAATGACTCTGAC 546
 Db 36 AATGATGATGAGTAAATATTTAATTTAATGAAGAAAGAAATTTATGATTAATGATC 95
 Oy 547 AAGTGAATTTCTACTTCGACTTGTCTAAGGCCAACTGAATGACGTTGGCCCAATGAC 606
 Db 96 GCAATAAATTTTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 155
 Oy 607 TATTCGAAATTCATTCATTTGAAATGATGACGCAACGAGTGTGACGCTATTGAAGAG 666
 Db 156 TATTCGAAATTCCTGACATCTTAAATTTAGTGAAGAAAGAAATGCTTAAAAA 215
 Oy 667 TTGTCCTTCGATATGCAAGCCTCTCGAACATCAAGACAAATGTGGAAAGATGAA 726
 Db 216 GTAATTTTGTATGATTAAGAAACCAATAGAAATATTCAGACGATATGAAAGTTAGA 275
 Oy 727 GATTATATTTAAAGATATAGAACCATCGAACATTTAAGAGCTGTGCAAGATCC 786
 Db 276 ATTTACATATGAAAGAAATTAAGAACTGTGACGCTTTAAAGCTTTTCTTCTGAA 335
 Oy 787 AAAAAAGACCTA-----GACAAAATTAAGATGCAACC 819
 Db 336 ACAAAAAAATATCAACCTTAAGTAACGAAGATTCATGACGCTAGTTGTATAGCCAT 395
 Oy 820 AAGAGGAGAAAGAAAGAAAGTTGTACAGGCCAGTACGACCTGTCCATATACAA 879
 Db 396 AATATTAATTAATAAAAAAACCATATATACCAAGTATGATCAATGATATATTTAA 455
 Oy 880 CAGCTTGAAGAACCCATTAACCTCACTGACGCTGACGATGCAAGCGCATTAACAC 939
 Db 456 CAATTAGCTGAATACAAAGGTTGCGAAGCTTTGAAAGAAAGCTTTCTCATTAAG 515
 Oy 940 AAGATGAATAATATCAAAAGACTGTCGACAGATTATGAATTAAGATCTCCGCCA 999
 Db 516 AAAATGATGCATCAACACATTATGCGAACAAATTTGAAGTTCTCATCTGCCCCGTC 575
 Oy 1000 GCCAAGCTGTGGACAC 1016
 Db 576 GTCACTGCCGGAACAC 592

RESULT 20
 A0086246 500 bp mRNA linear EST 27-JAN-2001
 LOCUS A0086246
 DEFINITION A0086246 Sugano Malaria cDNA library Plasmodium falci-parum 3D7 cDNA
 clone xPfn2175 similar to P.falci-parum gp190 (MSA), PMSA)
 for precursor of major merozoite surface antigens, mRNA sequence.
 ACCESSION A0086246.1 GI:12388387
 VERSION A0086246.1
 KEYWORDS EST.
 SOURCE Plasmodium falci-parum 3D7.
 ORGANISM Plasmodium falci-parum 3D7.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 500)
 AUTHORS Watanabe, J., Sasaki, M., Suzuki, Y. and Sugano, S.
 TITLE FULL-malaria: a database for a full-length enriched cDNA library
 from human malaria parasite, Plasmodium falci-parum
 Nucleic Acids Res. 29 (1), 70-71 (2001)
 JOURNAL MEDLINE 20574754
 COMMENT Contact: Junichi Watanabe
 Institute of Medical Science
 The University of Tokyo, Department of Parasitology
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanab@manage.ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1..500
/organism="Plasmodium falciparum 3D7"
/strain="3D7"
/db_xref="taxon:36329"
/clone="ypn2175"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 244 a 69 c 72 g 115 t
ORIGIN

Query Match 2.2%; Score 108.8; DB 9; Length 500;

Best Local Similarity 58.2%; Pred. No. 6e-14; Mismatches 137; Indels 0; Gaps 0;

Matches 191; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1745 AGACGCTGTGAGACATTAGAGAGATGAGAGACAGTTGTTGAGAGAGATTACAA 1804

DB 19 ATTCATTAATTTACAGAGAAATTAAGAGAGTGAACAATTTCTAGAAAAAATTTA 78

QY 1805 AAGACGAAATAAACAGATGAGAGATCCCTGGAGTCTCCGATTTGTTAAAGTCCAG 1864

DB 79 AAGGACTAACACATCAGCAATGTTCCCTAGAGATCTGATTTGTTAAATTCAGAG 138

QY 1865 TCGAGAGGTGCTCTCATTGAGACAGATGAGATCAAGAGTCACTCATCTGTA 1924

DB 139 TACAAAAAGTTTATTAATTTAAAAAATGAGACCTTAAGAAAGATGAAATTTTAA 198

QY 1925 AGAACCTGAGTTAAACATTAATATCATGTGCCGATATGTTTAAGCAGAGATTAAGC 1984

DB 199 AAATGACACACATTAAGATAGTATTCATGATACCAATATTTTAAACCAAAATTAAC 258

QY 1985 AGGACCATTAATCTCTATCTGATCTGACAGAGAGATAGCAAGAGTTCATGC 2044

DB 259 CAGAACCATATTTATTTAATTTGTTTAAAAAAGAGATAGTAAATTTAAAGATTTATAC 318

QY 2045 CCAAGTCGAGAGCCGATCAACGAGA 2072

DB 319 CAAAAGTAAAGACATGTTAAAGAGAGA 346

RESULT 21

BM159731

LOCUS BM159731 755 bp mRNA linear EST 04-DEC-2001

DEFINITION EST162224 PyBS Plasmodium yoelii yoelii cDNA clone pYCH95 5' end,

mRNA sequence.

ACCESSION BM159731

VERSION BM159731.1 GI:17305412

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Plasmodium yoelii yoelii.
Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 755)
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
Fraser, C.M. and Carucci, D.J.
Plasmodium yoelii EST project at TIGR
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208

Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1..755

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCH95"

/clone_lib="PyBS"

/dev_stage="Asexual blood stages"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cBYJ mice infected with PY17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidinium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

terminals were treated with Pfu DNA polymerase and EcoRI

adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybridZAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybridZAP vector and plasmid DNA

isolated."

BASE COUNT

323 a 103 c 136 g 193 t

ORIGIN

Query Match 2.2%; Score 108.8; DB 13; Length 755;

Best Local Similarity 52.4%; Pred. No. 6.6e-14; Mismatches 255; Conservative 0; Mismatches 232; Indels 27; Gaps 1;

Matches 285; Conservative 0; Mismatches 232; Indels 27; Gaps 1;

QY 487 AACATTCATGGCTTCAATTTCTGATTTGAGCTTACGAGATCAATGAATCCTGTAC 546

DB 39 AATGATATGAGTAAATTTAATTTAAGAAAGTAAAGAAATTTATGATTAATGATTAAC 98

QY 547 AACTGAATTTCTACTTCGACTTGCATAGGCGCAAACTGATATGCTTGGCATGAC 606

DB 99 GCAATTAATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 158

QY 607 TATTGCAATTTCCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 666

DB 159 TATTGGAATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 218

QY 667 TTGCTTTCGATATGCAAGCCCTTCGACACATCAAGGACATGTGGAAAGATGAA 726

DB 219 GTATTTTATGATTAAGAAACCAATAGAAATATTCAGACATATTTGAAGTTAGAA 278

QY 727 GATTATATTAAGATTAAGAGACCATGCAACATTTAGCAGCTGATCGAAGATCC 786

DB 279 ATTACATAGAAAGAAATTAAGAACTGTGACGTTTAAACCTCTTATTTGCTGAAG 338

QY 787 AAAAGACCATTA-----GACAAATTAAGATGCAAC 819

DB 339 ACATAAAATTAACACCTGAGGTAAACGAGATTTGCAATGACCTTGTGATAGCAT 398

QY 820 AAGGAGAAAGAAAGAAAGATTGTACAGCCAGTACAGCTGTCCATTAACAA 879

DB 399 AATATTAATTAAGAAACCAATATACAGATGATGATGATGATGATGATGATGAT 458

QY 880 CAGCTTGAAGAGCCCATTAACCTGATGATGATGATGATGATGATGATGATGAT 939

DB 459 CAATTAGCTGAATTAACAAAGGTTGCAAGCTTTGAAAGAGAGATTTCTACATTAAG 518

QY 940 AAGATGAATATTCAGAAAGATGCTGCAACATTAATTAAGATTCCTGCCCA 999

DB 519 AAAAATGATGCCATCAACCATTAATGCAACAAATGAAGTCTCAATGCTGCCCGCTC 578

QY 1000 GCCA 1003

DB 579 GTCA 582

```

RESULT 22
LOCUS      BM160423
DEFINITION EST562946 PyBS Plasmodium yoelii yoelii cDNA clone pYJCRII 5' end,
ACCESSION  BM160423
VERSION     BM160423.1
KEYWORDS   EST.
SOURCE      Plasmodium yoelii yoelii.
ORGANISM   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 753)
AUTHORS    Fraser,C.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
           Carlton,J.M. and Carucci,D.J.
TITLE      Plasmodium yoelii EST project at TIGR
JOURNAL    Unpublished (2001)
COMMENT    Contact: Jane Carlton
           Parasite Genomics Group
           The Institute for Genomic Research
           9712 Medical Center Drive, Rockville, MD 20850, USA
           Tel: 301-530-9319
           Fax: 301-838-0208
           Email: carlton@tigr.org
           For clone info, please contact the Malaria Research and Reference
           Reagent Resource Center, ATCC
           http://www.malaria.mr4.org/mr4pages/index.html
           Seq primer: ADF

FEATURES
Source
Location/Qualifiers
1..753
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="pYJCRII"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cBYJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dt)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT      335 a      101 c      113 g      204 t
ORIGIN
Query Match      2.2%; Score 107.6; DB 13; Length 753;
Best Local Similarity 58.9%; Pred. No. 1.2e-13;
Matches 185; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 487 AACATTCATGCGCTCAATATCTGAGCTTACGAGATCAATGACTCCTGAC 546
    ||| |||| ||||| ||||| ||| ||||| ||||| ||||| |||||
DB 172 AATGATATGAGTAATAATATTATTAAGTAAGAATTAATGATTAATGCAC 231
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 547 AAGTTGAAATTTCTACTGACTGCTAAGGCCAAAGCAATGACGTTGCCCAATGAC 606
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 232 GCAATAAATTTTATTAATGATGCTAGAGTAATAATTAATGATATCTGCAAAATAAT 291
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 607 TATGTCAATTCATTCATTCATTTGAGATCAGACCAAGCGTTGAGCTATGGAAGAG 666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 292 TATTTGTGAATTCCTGAAACATCTTAATAATTAGTGAAGAAGAAACGAATGCTTAAGAAA 351
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 667 TTGCTCTCGATATGCGCAAGCCTCTGACACATCAAGACAAATGCGGAAGATGAA 726
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 352 GTAATTTTGGGTTATATGAAAACCAATAGAAAATATTCAAGACGATATTAAAAAGTTAGAA 411
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 727 GATTTATTTTAAAGATTAAGAACCATCGAAGACATTAAAGCTGATCGAAGATCC 786
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 412 ATTTACATGAAAGAAATTAAGAAACTGTTGACGCTTAAACGCTCTATTCGTGAAGAA 471
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 787 AAAAGACCATATGA 800
    ||| |||||
DB 472 ACAAATAAATAATACA 485
    ||| |||||

RESULT 23
LOCUS      BM165725
DEFINITION EST568248 PyBS Plasmodium yoelii yoelii cDNA clone pYCMR8 5' end,
ACCESSION  BM165725
VERSION     BM165725.1
KEYWORDS   EST.
SOURCE      Plasmodium yoelii yoelii.
ORGANISM   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 678)
AUTHORS    Fraser,C.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
           Carlton,J.M. and Carucci,D.J.
TITLE      Plasmodium yoelii EST project at TIGR
JOURNAL    Unpublished (2001)
COMMENT    Contact: Jane Carlton
           Parasite Genomics Group
           The Institute for Genomic Research
           9712 Medical Center Drive, Rockville, MD 20850, USA
           Tel: 301-530-9319
           Fax: 301-838-0208
           Email: carlton@tigr.org
           For clone info, please contact the Malaria Research and Reference
           Reagent Resource Center, ATCC
           http://www.malaria.mr4.org/mr4pages/index.html
           Seq primer: ADF

FEATURES
Source
Location/Qualifiers
1..678
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="pYCMR78"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cBYJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dt)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT      291 a      93 c      120 g      174 t
ORIGIN
Query Match      2.1%; Score 105; DB 13; Length 678;
Best Local Similarity 52.4%; Pred. No. 4.7e-13;
Matches 277; Conservative 0; Mismatches 225; Indels 27; Gaps 1;

```

Query Match	Best Local Similarity	Score	DB	Length
Matches 183; Conservative	57.7%	102.6	DB 13	402
	0; Mismatches 134; Indels	0; Gaps		0;
QY 3923	TCGACCTTCACATGTCACCTGGAAGACATTCGTGAAACGCGCTTTATATAGAGAGAAATTT	3982		
Db 5	TTACATTTTAATTTAATTTTGAACGATATCTTAATTCACGCTTTTGAAGAACCAAAATATTT	64		
QY 3983	TCGAGAACGCTCTTGAGAGAGCGACTGTTGATTCCTTATTAAGACCTGACCTCTTAACTACG	4042		
Db 65	TCTTAGATGATTTAGAAATCTGATTTAAAGCAATTTAAACATATATTCCTCAATTAATGATACA	124		
QY 4043	TTGTCAGACGACCATACACAGTTCCTCCATTAAGAGAGAGAGGATTAATTTCTGCTAGTT	4102		
Db 125	TTATTTGAAGATTCATTTTAATTTATTTGAATTCGAGACAAAACAAACACACTTTTAAAAAGTT	184		
QY 4103	ACAACATATATCAGAGACTTCATCGACACCGATATATCAATTTGCTATAGATGCTGGGGT	4162		
Db 185	ACAATATATATAAAGAAATACAGTAAATGATATTAATTTCCACAGAGGATTAATAGTT	244		
QY 4163	ATTACAAATATCCTGAGCGCAAAATATCAAGTCTGACCTGACCTCTTAAAAAGTATATCA	4222		
Db 245	ATTATGAAAGAGTTTATACGAAATATTAAGATGATTTGAATCAATTTAAAAAGTTATCA	304		
QY 4223	ACGATTAAGCAGCGCGAG	4239		
Db 305	AAGAAGAAAGAGAGAG	321		
RESULT 25				
LOCUS	N97689	400 bp	mRNA	linear
DEFINITION	1114C3 czappedd3.1, Debopam Chakrabarti Plasmodium falciparum cDNA			
ACCESSION	N97689			
VERSION	N97689.1	GI:1674723		
KEYWORDS	EST.			
ORGANISM	Plasmodium falciparum			
SOURCE	Enkaryoto; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.			
REFERENCE	1 (bases 1 to 400)			
AUTHORS	Dame,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z., Coppel,R., Coman,A., Craig,A., Fischer,K., Foster,J., Goodman,N., Hunteberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A., Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su,X.-Z., Thompson,J.K., Vitali,F., Wellens,T.E. and Werner,E.			
TITLE	Current status of the Plasmodium falciparum genome project			
JOURNAL	Mol. Biochem. Parasitol.	79, 1-12		(1996)
MEDLINE	97001675			
COMMENT	Contact: Debopam Chakrabarti			

FEATURES

source

Location/Qualifiers

1. 647

/organism="Plasmodium berghei"

/strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"

/db_xref="taxon:3821"

/clone_lib="Pb MBN #21"

/dev_stage="asexual blood forms"

/lab_host="Mus musculus"

/note="Vector: pBluescript SK(+) vector DNA, phagemid

excised from lambda ZAP, Site_1: EcoRV, Site_2: EcoRV;

Genomic DNA was prepared from asynchronous blood stage

forms of the cloned ANKA isolate of P. berghei grown in

laboratory Swiss white mice. The DNA was purified from

contaminating host DNA by Hoechst Dye 33258-CsCl

ultracentrifugation and precipitated. Purified DNA was

digested with mung bean nuclease in the presence of 36-38%

formamide at 50 C, as described (Vernick, K.D., Impericki,

R.B., and McCutchan, T.F., 1988. Nucleic Acids Research

16:6883-6896). The ends of the digestion fragments were

polished using T4 DNA polymerase, and the fragments size

selected in the range 500-2000 bp. These were ligated into

the EcoRV-cleaved and dephosphorylated pBluescript SK(+)

vector. Recombinant plasmids were used to transform E.

coli XL10-Gold host cells."

BASE COUNT 282 a 70 c 92 g 202 t 1 others

ORIGIN

Query Match

Best Local Similarity 56.9%; Pred. No. 2.8e-11;

Matches 177; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Score 97.2; DB 17; Length 647;

2.0%;

487 AACATTCAGCGCTCAATATGATGATGACGTTAGACAGATCAATGATCCTGTAC 546

316 AATTAATATGACGCTCAATATGATGATGACGTTAGACAGATCAATGATCCTGTAC 375

547 AATTAATATGACGCTCAATATGATGATGACGTTAGACAGATCAATGATCCTGTAC 606

376 GCATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 435

607 TATTCGAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 666

436 TATTCGAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 495

667 TTGGCTCTCGGATATGCGAAGCCTCTCGACACATCAAGACATGTGGGAAGATGGA 726

496 GTAATTTAGGTATAGAAACCAATAGAAATATTTCAAGCATTTTGAAGAATTGAA 555

727 GATTATATTAAGAAATAGAAACCAATAGAAATATTTCAAGCATTTTGAAGAATTGAA 786

556 ATTACATTAACAAATAAAGAAACCTTTACAGCTTTAAACACTCTTATTTGTTGAAGA 615

787 AAAAGACCAT 797

616 AACAAAAAAT 626

RESULT 28

BM159472

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Jane Carlton

EST561995 PyBS Plasmodium yoelii yoelii CDNA clone pYCJ56 5' end,

660 bp mRNA linear EST 04-DEC-2001

BM159472 GI:17305153

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.orgFor clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADP.

Location/Qualifiers

1. 660

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone_lib="PyBS"

/dev_stage="asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cByJ mice infected with Py17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidinium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

terminals were treated with Pfu DNA polymerase and EcoRI

adaptors. Ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybridZAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybridZAP vector and plasmid DNA

isolated."

BASE COUNT 281 a 68 c 95 g 216 t

ORIGIN

Query Match

Best Local Similarity 60.5%; Pred. No. 6.7e-10;

Matches 150; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Score 91.2; DB 13; Length 660;

1.8%;

487 AACATTCAGCGCTCAATATGATGATGACGTTAGACAGATCAATGATCCTGTAC 546

411 AATGATATGACGCTCAATATGATGATGACGTTAGACAGATCAATGATCCTGTAC 470

547 AATGATATGACGCTCAATATGATGATGACGTTAGACAGATCAATGATCCTGTAC 606

471 GCATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 530

607 TATTCGAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 666

531 TATTCGAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 590

667 TTGGCTCTCGGATATGCGAAGCCTCTCGACACATCAAGACATGTGGGAAGATGGA 726

591 GTAATTTAGGTATAGAAACCAATAGAAATATTTCAAGCATTTTGAAGAATTGAA 650

727 GATTATAT 734

651 ATTACAT 658

RESULT 29

BM170928

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Plasmodium yoelii yoelii.

777 bp mRNA linear EST 04-DEC-2001

EST573451 PyBS Plasmodium yoelii yoelii CDNA clone pYCON24 5' end,

BM170928 GI:17304160

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii.

Query Match	Best Local Similarity	Score	DB	Length
Matches 115; Conservative	70.6%;	0; Mismatches	48; Indels	0; Gaps
0y 4777	GAAGACAGCGGCTTAAGGAAAGAAATTCACATCGAGTGTACTAAGCCGACATCTAT	4836		
Db 2	GAGATTGACGTAGCAGCAGAGAAAGAAATTCACATGTGATGTACTAAGCTGATTTAT	61		
0y 4837	CCACTCTTCGAGCGGAGTTTTTGTCTCAAGCTTAATTTCTGGGCAATCTCTCTGCG	4896		
Db 62	CCACTTTTCGATGTGATTTTCTGACAGTTCCTACTACTCTTGAAGAATCATCTTATTA	121		
0y 4897	ATCCATCATGCTGATCCTCTACAGCTTCATCTAATGATCATGATG	4939		
Db 122	ATACATCATGTTAATTAATTAATACAGTTTCATTTAAAAAATGTAAG	164		
RESULT 33				
LOCUS	B0577302	352 bp	mRNA	linear
DEFINITION	PfEST0ab17c05.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9VA18 Q9VA18 SURFACE PROTEIN-1; mRNA sequence.			
ACCESSION	B0577302			
VERSION	B0577302.1	GI:21480619		
KEYWORDS	EST.			
SOURCE	Malaria parasite P. falciparum.			
ORGANISM	Plasmodium falciparum			
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS	1 (bases 1 to 352)			
	Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagaris, V., R., Belagorod, L., Franklin, C., Carr, L., Grow, A., Mauter, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.			
TITLE	WashU Plasmodium EST Project			
JOURNAL	unpublished (2001)			
COMMENT	Contact: L. David Sibley WashU Plasmodium EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University			

FEATURES	Seq primer: -40Up from GlbCO.
Location/Qualifiers	
1..352	
/organism="Plasmodium falciparum"	
/db_xref="taxon:5833"	
/clone_lib="plasmodium falciparum 3D7 asexual cDNA"	
/lab_host="DH10B (Geneshop, Invitrogen, Inc.)"	
/note="Vector: pBluescript SK plus; Site_1: EcoRI. Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage zaponin (0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin MagneSphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 zapII vector using the zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the Exassist helper phage (Stratagene), the phagemids were preplated with peg 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."	
BASE COUNT	143 a 33 c 32 g 143 t 1 others
ORIGIN	
Query Match	1.7%; Score 85.2; DB 14; Length 352;
Best Local Similarity	70.4%; Pred. No. 1.4e-08;
Matches 114; Conservative 0; Mismatches 48; Indels 0; Gaps 0;	
QY 4778 AAGACAGCGCTCTAACGGAAGAAATACATCGAGTCTACTAGCCGAGCTCTATC	4837
Db 1 AAGATTGAGTACAGCAGCAAGAAATACATGTGATGTACATCACTGATTTATC	60
QY 4838 CACCTTCGAGGGGATTTTTCGCCAGCTCAATTTCCGGGATCTCCTCCGGCGA	4897
Db 61 CACCTTCGAGGGATTTTTCGCCAGCTCTTAACCTTTAGGAATATCATTTATTA	120
QY 4898 TCTCTATGCTGATCTCTACAGCTTCACTCATCTAATGATCGATG	4939
Db 121 TACTCATGTTAATATTATACAGTTTCATTTAATAAATGTTAG	162
RESULT 34	
T18003	282 bp mRNA linear EST 30-AUG-1994
LOCUS	0325c3 cbSPFHB3.1, Debopam Chakrabarti Plasmodium falciparum cDNA
DEFINITION	clone 0325c 5', mRNA sequence.
ACCESSION	T18003
VERSION	T18003.1
KEYWORDS	GI:462789
SOURCE	EST.
ORGANISM	malaria parasite P. falciparum.
REFERENCE	Plasmodium falciparum
AUTHORS	Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
TITLE	1 (bases 1 to 282)
JOURNAL	Chakrabarti, D., Reddy, G.R., Dame, J.B., Almtira, E.C., Lalpis, P.J.,
MEDLINE	Perl, R.J., Yang, T.P., Rowe, T.C. and Schuster, S.M.
COMMENT	Analysis of Expressed Sequence Tags from Plasmodium falciparum
	Mol. Biochem. Parasitol. 66, 97-104 (1994)
	95075403
	Contact: Debopam Chakrabarti
	Department of Molecular Biology and Microbiology
	University of Central Florida
	Orlando, FL 32816-2360
	Tel: 407 384 2061
	Fax: 407 384 3095
	Email: dchak@pegasus.cc.ucf.edu
	Seq primer: T3.
	Location/Qualifiers
FEATURES	
source	1..282
	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/clone="0325c"

FEATURES	Seq primer: -40Up from GlbCO.
Location/Qualifiers	
1..352	
/organism="Plasmodium falciparum"	
/db_xref="taxon:5833"	
/clone_lib="plasmodium falciparum 3D7 asexual cDNA"	
/lab_host="DH10B (Geneshop, Invitrogen, Inc.)"	
/note="Vector: pBluescript SK plus; Site_1: EcoRI. Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage zaponin (0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin MagneSphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 zapII vector using the zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the Exassist helper phage (Stratagene), the phagemids were preplated with peg 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."	
BASE COUNT	143 a 33 c 32 g 143 t 1 others
ORIGIN	
Query Match	1.7%; Score 85.2; DB 14; Length 352;
Best Local Similarity	70.4%; Pred. No. 1.4e-08;
Matches 114; Conservative 0; Mismatches 48; Indels 0; Gaps 0;	
QY 4778 AAGACAGCGCTCTAACGGAAGAAATACATCGAGTCTACTAGCCGAGCTCTATC	4837
Db 1 AAGATTGAGTACAGCAGCAAGAAATACATGTGATGTACATCACTGATTTATC	60
QY 4838 CACCTTCGAGGGGATTTTTCGCCAGCTCAATTTCCGGGATCTCCTCCGGCGA	4897
Db 61 CACCTTCGAGGGATTTTTCGCCAGCTCTTAACCTTTAGGAATATCATTTATTA	120
QY 4898 TCTCTATGCTGATCTCTACAGCTTCACTCATCTAATGATCGATG	4939
Db 121 TACTCATGTTAATATTATACAGTTTCATTTAATAAATGTTAG	162
RESULT 34	
T18003	282 bp mRNA linear EST 30-AUG-1994
LOCUS	0325c3 cbSPFHB3.1, Debopam Chakrabarti Plasmodium falciparum cDNA
DEFINITION	clone 0325c 5', mRNA sequence.
ACCESSION	T18003
VERSION	T18003.1
KEYWORDS	GI:462789
SOURCE	EST.
ORGANISM	malaria parasite P. falciparum.
REFERENCE	Plasmodium falciparum
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
TITLE	1 (bases 1 to 282)
JOURNAL	Chakrabarti, D., Reddy, G.R., Dame, J.B., Almtira, E.C., Lalpis, P.J.,
MEDLINE	Perl, R.J., Yang, T.P., Rowe, T.C. and Schuster, S.M.
COMMENT	Analysis of Expressed Sequence Tags from Plasmodium falciparum
	Mol. Biochem. Parasitol. 66, 97-104 (1994)
	95075403
	Contact: Debopam Chakrabarti
	Department of Molecular Biology and Microbiology
	University of Central Florida
	Orlando, FL 32816-2360
	Tel: 407 384 2061
	Fax: 407 384 3095
	Email: dchak@pegasus.cc.ucf.edu
	Seq primer: T3.
	Location/Qualifiers
FEATURES	
source	1..282
	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/clone="0325c"

QY 953 TCAGAACTGCTGCAGCAGATTAAATGAATTAAGATTCCTCCGACCA 1003
|||||
Db 421 TCAGAACTGCTGCAGCAGATTAAATGAATTCCTCCGACCA 471

RESULT 38
BM166688
LOCUS BM166688 769 bp mRNA linear EST 04-DEC-2001
DEFINITION EST169211 PYBS Plasmodium yoellii yoellii cDNA clone pYCNV80 5' end,
mRNA sequence.
BM166688
ACCESSION BM166688.1 GI:17299920
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.m4.org/mr4pages/index.html
Seq primer: ADF.

Location/Qualifiers
1..769
/organism="Plasmodium yoellii yoellii"
/strain="17YL"
/db_xref="taxon:73239"
/clone_11b="PYBS"
/clone_11b="pYCNV80"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Pyl17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT 340 a 101 c 104 g 224 t
ORIGIN

Query Match 1.6%; Score 79; DB 13; Length 769;
Best Local Similarity 46.6%; Pred. No. 4.2e-07;
Matches 285; Conservative 0; Mismatches 325; Indels 1; Gaps 1;

QY 2908 TTCACCTACTCTGCGTAATTAAGCCGATATCACTCTTTAAGCATGATCTAA 2967
|||||
Db 111 TTCATTAATTTTATTAATAAATGAATTAACAAGATCTGACACCTCGAAAA 170

QY 2968 CGTAAGACCTGGAAGAGCATCATTAAGCTGAAGAGACACTGACCTGCAGC 3027
|||||
Db 171 GTTAATCATATTATTTCTGAATCGCTCACTTAAGGAATTAACAGAACATTATATGAT 230

QY 3028 CTGTACACAGATAGCAAACTGAAACTGAGAGACTCTTGACAGAGAAGACATCGCC 3087
|||||
Db 231 CGTTATTTACATATATTAATTAATTAAGAAAGATTATTAACAACATGAACAAATTCAA 290

QY 3088 AAGTATTAAGATGAGATCAAGAAAGTGTGACTCTCTCAAGAGAGCTTGAAGCAATC 3147
|||||
Db 291 CTAAACATGACAAATTAATTAAGATCTTAGATTAATTAAGCAATTAATTAAGAAAA 350

QY 3148 AACTCAGTACAAATTCGAAACACGATCTGACAGAACTTCTCACTGTTTCAACAGAA 3207
|||||
Db 351 CAACCTTAATGCGCTATTTTATATTAATTAAGTTATGTAATTTCTTTAACAAGAGA 410

QY 3208 AAGGAAGCCGAGATCGCCGAGACAGAAACACTCTGGAGAACACCAAGATTCTTCAAA 3267
|||||
Db 411 AGAAGAGCTGAAACAAATATGATGTAATGCAATTAATTAATTAATTAATTAATTA 470

QY 3268 CACTACAAAGGCGCTGCAAGTATTAATGCGAGCTCTTCCTGAAAGACTCTCTCC 3327
|||||
Db 471 TACTACAAAGCTGCTGACTAAATTTTACTCTGAAAGCTGTTCTTTAACAACATTATCT 530

QY 3328 GAGCAGAGCATCCAGACCGAGATTAACCTACGCCAGCTCGAGAACTTCAAGCTCTGCT 3387
|||||
Db 531 AAGCATCACTTACAGAGAAATCCATTATTTGAATATGAAATTCAGAGCATACAGT 590

QY 3388 AAGCTGGAAGGCAAGCTGAAGCAACCTGAACTGGAAGAGAGAGCTCAGCTACCT 3446
|||||
Db 591 CGATTGAATTAAGATTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 650

QY 3447 CTTGAGGAGACGATCAGCTGATGCTGCGAGCTCAAGGAAGTATTAAGAAAGAACTA 3506
|||||
Db 651 ATGAGAGGTTTACACGAGTATTAAGAAATTTAAAGAACTTAAGATTAAGACTA 710

QY 3507 CACCGCAATA 3517
|||||
Db 711 TACCGAAAAA 721

RESULT 39
B13042 1147 bp DNA linear GSS 14-MAY-1997
LOCUS B13042/c T30M24-Sp6.1 TAMU Arabidopsis thaliana genomic clone T30M24, DNA
DEFINITION
sequence.
B13042.1 GI:2094174
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

thale cress.
Arabidopsis thaliana
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
Rosidae: eurosids II: Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1147)
Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.

TITLE
JOURNAL
COMMENT

BAC End Sequences at ATGC
Unpublished (1997)
Other_GSSs: T30M24-Sp6, T30M24-Sp6.2
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends
High quality sequence start: 289
High quality sequence stop: 292.
Location/Qualifiers
1..1147
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"


```

      /clone="T30M24"
      /sex="hermaphrodite"
      /note="Vector: Belobacii; Site_1: HindIII; Site_2: HindIII
      ; Produced by Rod Wing"
BASE COUNT      3 a      4 c      46 g      553 t      541 others
ORIGIN

```

```

Query Match      1.6%: Score 78.8; DB 17; Length 1147;
Best Local Similarity 23.8%: Pred. No. 5.2e-07;
Matches 251; Conservative 0; Mismatches 805; Indels 0; Gaps 0;

Qy 695 ACAACATCAAGGACATGTTGGGAAAGATGAAATTTATTTAAAGATTAAGACCA 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1145 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1086

Qy 755 TCGAGACATTACGAGCTGATCGAAGATCCAAAAAGACCTAGACAAAATTAAAGTG 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1085 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1026

Qy 815 CAACCAAGAGGAGAAAGAAAGAGTGTACCGCCAGACGACCTGTCATCTATA 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1025 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 966

Qy 875 ACAACACCTTGAAGAGCCCATTAACCTCATCAGCTACTGGAAGCGCATAGACACC 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 NANANNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNAN 906

Qy 935 TCAAGACAAATGAAATTTCAAGAGCTGCTCGACAACTTATGAAATTAAGATCCTC 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 846

Qy 995 CGCCGACACTCTGGAACACCCCTTAACAGCTCTGAGAGAGAAAGATAGAGG 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 845 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 786

Qy 1055 AGCAGCAAGAGAGATCAAGAGATCGCCAAACCATTAAGTCAACATAGATTCTCT 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 NNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 726

Qy 1115 TTACTGATCCCTTGAGCTGAGTACTTCTTGAGAGAGAAATTAAGATATAGACATCT 1174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 NAAANAAANAAANAAANAAANAAANAAANAAANAAANAAANAAANAAAN 666

Qy 1175 CCGCCAAAGTCGAGCAAGGAATCAACGAACTAATGAATATCCATGGTGTGAGCT 1234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 606

Qy 1235 ACCCTCTCTTATTAACGATATCAACAGCTCTCAAGAGCTCAATAGCTCGGTGACT 1294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 ANNNNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 546

Qy 1295 TGATTTAACCCCTTGATTTATAGAAAGAACCTTAAGATATCTACACAGCAATGAGA 1354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 486

Qy 1355 GAAAGAACTTTATCAACGAATCAAGGAGAGATCAAAATGGAAGAAATTTGAGA 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 426

Qy 1415 GTGACAGAAAGATTACGAGACGCGACAAAGCTTAAGCATTTCACTAAAGATGTG 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 366

Qy 1475 AAAAGCTGCTGAAGAGATCTATGATTCAAATTTCAACATTAACATGACCTGACCACT 1534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 NNNNANANANNNANNNANNNANNNANNNANNNANNNANNNANNNNNNNNN 306

Qy 1535 TCGAGAAATGATGGGAAAGCGTACTCTTACAAAGTGAGAGAACTGACACACCTAATA 1594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 246

Qy 1595 CCTTTCATCTATGAGAAATTTAAGCATTAATCTTGAGAGCTCACCAAGCTCTTAAGT 1654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 245 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 186
Qy 1655 ATATGAGAGACTATTTCTGCGGAACATTTGTTGTGAGAAAGAACTAAAGATTACAAGA 1714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 126
Qy 1715 ATTCATAGTAAGATGAGAAACGAGATCGAGACC 1750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 90

```

```

RESULT 40
BM656118      503 bp      mRNA      linear      EST 26-FEB-2002
LOCUS
DEFINITION
17000687387385 A.Gam.ad.cdNAL Anopheles gambiae cdNA clone
19600449658190 5', mRNA sequence.
BM656118
ACCESSION
BM656118.1 GI:18955629
VERSION
KEYWORDS
EST.
SOURCE
African malaria mosquito.
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE
1 (bases 1 to 503)
Holt R.A., Jin J.-J., Murphy S.D., Evans C.A., Kraft C.L., Charlab
R., Collins F.H., Venter J.C., and Hoffman S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltR@celera.com
Plate: ND01004HNK row: K column: 20
Seq primer: M13 Reverse.

```

```

FEATURES
    source
        1..503
            location/qualifiers
                /organism="Anopheles gambiae"
                /strain="RSP-ST (Reduced susc. to Permethrin - std.
                chromosome)"
                /db_xref="taxon:7165"
                /clone="19600449658190"
                /clone_lib="A.Gam.ad.cdNAL"
                /dev_stage="Adult"
                /lab_host="DH10b"
                /note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
                adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                cDNA inserts >500 bp cloned directionally into pSport 1.
                Not I site is 3'. Clones available through the Malaria
                Research and Reference Reagent Resource Center
                (www.malaria.mr4.org)."

```

```

BASE COUNT      300 a      124 c      72 g      7 t
ORIGIN
Query Match      1.6%: Score 77.2; DB 13; Length 503;
Best Local Similarity 51.0%: Pred. No. 9.9e-07;
Matches 209; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

```

```

Qy 693 CGACAACTCAAGGACATGTGGGAAAGATGAAATTTATTTAAAGATTAAGAACAC 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 148

Qy 753 CATCGAACAATTAAAGCATGATCGAAGATCCAAAAAGACCAATAGCAAAAAATTAAGAA 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 208

Qy 813 TGCAACCAAGGAGAGAAAGAAAGATTTGACAGGCCCAAGTACGACTGTCCATCTA 872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 CAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 268

```

QY 873 TACAAACGCTTGAGAGAGCCATACCTACGCTAGTGAGAGAGCGCTAGACAC 932
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 Db 269 GAACAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAAC 328
 QY 933 CCTCAAGAAAGATGAATATATCAAGAACTGCTCGACAAAGATTATGAATTAAGATCC 992
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 Db 329 GAACAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAAC 385
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 QY 993 TCCGACGACCACTCTGGAGACACCCCTACACGCTGCTGGACAGACAGAAATAGA 1052
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 Db 386 CAAGAACAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAA 445
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 QY 1053 GGAGCAGAGAAAGATCAAGAGATCGCCCAAAACCAATTAGTCAACA 1102
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 Db 446 CAAGAACAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAG 495
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -

RESULT 41
 BM159366 344 bp mRNA linear EST 04-DEC-2001
 LOCUS BM159366
 DEFINITION EST561889 PyBS Plasmodium yoelii yoelii cDNA clone PYCJD17 5' end,
 mRNA sequence.
 ACCESSION BM159366.1 GI:17305047
 VERSION EST.
 KEYWORDS Plasmodium yoelii yoelii.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 344)
 AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
 Frazer, C.M., and Carucci, D.D.
 JOURNAL Plasmodium yoelii EST project at TIGR
 COMMENT Unpublished (2001)
 Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-530-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADP.
 Location/Qualifiers
 1..344
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCJD17"
 /clone_1lb="PyBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with PY17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase, and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (PAD-GAL4) was
 excised from the HybridZAP vector and plasmid DNA
 isolated."

BASE COUNT
 ORIGIN

145 a 65 c 59 g 75 t

Query Match 1.5%; Score 74.4; DB 13; Length 344;
 Best Local Similarity 57.9%; Pred. No. 3.9e-06;
 Matches 132; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1861 CAAAGTCAGAAAGTGTCTCTCATGAGCAAGATTGATGAACTCAAGAAAGCTCAATTCATT 1920
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 Db 2 CAAATCAAAAATCTTTATTTAAACAAACAAATTTGACCAATTAATTAACAGAGATGCT 61
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 QY 1921 CTGCAAGCGTGAGGTTAAACATTAATATATACATGCTCCGAATAGTATTAAGCAGAGAAAT 1980
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 Db 62 TTAAACAAGCTCAATTTAAAGCAACATATATGTTCCAAAAACATACGATATGAAAGA 121
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 QY 1981 AAGCAGAAACCATCTACTCATCTGCTACTGACAGAAAGATGACAACTGAAGATGTTCC 2040
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 Db 122 AAACCGAAGACATCATCTATTATTAATAGCTGTAAAAAAAGAGTTGACGACTGCCCAATTT 181
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 QY 2041 ATGCCCAAGTCGAGAGCCGTGATCAACGAAGAGAACAGACATTTAA 2088
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
 Db 182 ATTCCAAAATCGAAAGTATGATTGCTTAAAGAGAGAAAGAAATGGA 229
 ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -

RESULT 42
 A2813205 540 bp DNA linear GSS 20-FEB-2001
 LOCUS A2813205
 DEFINITION 2M0080C24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC2M0080C24 F, DNA sequence.
 ACCESSION A2813205
 VERSION A2813205.1 GI:12983113
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 540)
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,
 and Wright, D., Weiss, R.
 JOURNAL Mouse whole genome scaffolding with paired end reads from 10kb
 COMMENT Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0080 row: C column: 24
 Seq primer: CGTTGTAAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 540.
 Location/Qualifiers
 1..540
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0080C24"
 /clone_1lb="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus (C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to

FEATURES

source

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (gll4732114[plara129072.1], a copy-number inducible derivative of plasmid R1). The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT	284 a	54 c	151 g	51 t
ORIGIN				

Query Match	1.5%;	Score 73.8;	DB 17;	Length 540;
Best Local Similarity	49.9%;	Pred. No. 6e-06;		
Matches 186;	Conservative 0;	Mismatches 187;	Indels 0;	Gaps 0;

Oy	731	ATTTAAAAAGATTAAGAACCATCGAGACCTTTACGACGCTGATCGAAGATCCAAA	790
Db	115	AGATGAAGAAAGAACAAAGAACAGAACGAAGAAAGAAAGAAAGAAAGAAAGAA	174
Oy	791	AGACCATAGACAAAATAATTAAGATGCAACCAAGAGAGAAATAAAGAAAGTTGTACAGG	850
Db	175	AGAAAGAAAGAAAGACCTCGAAGAGAGAGAGAGAGCTAGAAGAAAGAAAGAGACTGGAAGA	234
Oy	851	CCGACGACGACCTGTCCATCTATTAACAAACAGCTTGAAAGACCCTTAACCTCATCAGCG	910
Db	235	AGAAAGAAAGAAAGAACAGAACAGAACAGAACGACTGAGAAAGAAAGAAAGAAAGAA	294
Oy	911	TACTGGAAAGCGCATGAGACACCCCTCAAGAGAGATTAATAATTCAAGAGACTGCTGCACA	970
Db	295	AGAAAGAAAGAAAGAACAAAGAACAGAACAGAACGAGAAAGAGAGACTGAGAAAGGAGAA	354
Oy	971	AGATTAAATGAATTAAGCAATCTCCGCCAGCCACACTCTGGAGAACCCCTTAACACGCTGC	1030
Db	355	AGAAAGACTGGAGAGAGAACAGACCTGAGAGAGAGAGAAAGAACAAAGAACAAAGAAC	414
Oy	1031	TGGACAGAACACAGAAAGATGAGAGACACAGAGAAAGATCAAGAGATTCGCCAAACCA	1090
Db	415	AGAACAGAAACAAAGAGAACAGAACAAAGAGAAAGAAAGAACAAACATCAAAAAGAA	474
Oy	1091	TTTAAGTTCAACAT	1103
Db	475	GCCAGGACAAATAT	487

RESULT 43	BM415494/c	948 bp	mRNA	linear	EST 28-JAN-2002
LOCUS	BM415494				
DEFINITION	OP0570 Mixed Stage EST's from <i>Globodera pallida</i> , the potato cyst nematode <i>Globodera pallida</i> cDNA, mRNA sequence.				

REFERENCE	1 (pages 1 to 948)
AUTHORS	Heer, J., Sosninski, B., Pokrzywa, R. M., Wiatry, A. and Opperman, C.
TITLE	Mixed Stage Ery's from Globodera pallida, the potato cyst nematode
JOURNAL	Unpublished (2001)
COMMENT	Contact: Opperman, C

FEATURES
source

```
/organism="Globodera pallida"  
/db_xref="taxon:36090"
```

```

clone_1st="mixed stage Est's from Globodera pallida, the
potato cyst nematode"
/vector: lambda GT11. This is a collaborative effort
between IACR-Rothamsted and North Carolina State
University. The library was constructed from mixed
stage G. pallida in lambda GT11 by Paul Burroughs,
IACR-Rothamsted."

```

BASE COUNT	16 a	89 c	20 g	744 t	79 others
ORIGIN					

Query Match	1.58;	Score 73.4;	DB 13;	Length 948;
Best Local Similarity	40.38;	Pred. No. 8.4e-06;		
Matches 363; Conservative	0;	Mismatches 532;	Indels 6;	Gaps 1;

[illegible]

OY 1594 A 1594
Db 26 A 26

RESULT 44
BE420745
LOCUS
DEFINITION HM002.B02 ITRC HMM Barley Leaf Library Hordum vulgare cDNA clone
HM002.B02, mRNA sequence.
BE420745
ACCESSION BE420745.1 GI:9418588
VERSION
KEYWORDS EST.
SOURCE Hordum vulgare.
ORGANISM Hordum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordum.
1 (bases 1 to 1885)
REFERENCE
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Landridge,P., Lazo,G.R., Lin,J.J., McStyre,P., Ogihara,Y.,
Pechioni,N., Qualset,C., Schuch,W., Selayrat,G., Shariflou,M.,
Sorrells,M., Warburton,M., and Wenzel,G.
International Triticeae EST Cooperative (ITREC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
JOURNAL
COMMENT Contact: Herrmann RG
Botanisches Institut der LMU
Menzinger Str. 67, D-80638 Muenchen GERMANY
Fax: 49 30 171683
Email: herrmann@botanik.biologie.uni-muenchen.de
International Triticeae EST Cooperative (ITREC)
http://heat.pw.usda.gov/genome.
Location/Qualifiers
FEATURES
source 1..1885
/organism="Hordum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HM002.B02"
/clone_lib="ITREC HMM Barley Leaf Library"
/tissue_type="leaf"
/dev_stage="14 day old"
/note="Vector: pBluescriptSK(-); 850 bp average insert
size."
BASE COUNT 1138 a 219 c 212 g 176 t 140 others
ORIGIN

Query Match 1.4%; Score 71.6; DB 10; Length 1885;
Best Local Similarity 37.6%; Pred. No. 2.6e-05;
Matches 362; Conservative 0; Mismatches 602; Indels 0; Gaps 0;

OY 631 AGATGAGAGCCACGAGTGGACGATTTGAAGAAGTGTCTTCGATATCGCAGCCT 690
Db 892 AA 951
OY 691 CTCGACACATCAAGCAATGTGGGAAGATGAGATTAATTAAGAATAAGAG 750
Db 952 NNN 1011
OY 751 ACCATGAGAACATTAAGAGTGTGAGAAATCCAAAAAGACATAGACAAATAAG 810
Db 1012 AA 1071
OY 811 AATGCAACAGAGAGAGAGAGAGAGAGTGTACAGGCCCAAGTACGCTGTCCATC 870
Db 1072 AA 1131
OY 871 TATAACAACAGCTGTAAGAGCCATTAACCTCATCGCTACTGGAAGCCCATAGAC 930
Db 1132 AA 1191

OY 931 ACCCTCAGAGAAGTGAATAATATCAAGAAGCTCGACAGATTAATGAATTAAGAT 990
Db 1192 AA 1251
OY 991 CCTCCGCCAGCCCACTCTGGGAACACCCCTTACACGCTGCTGGACAGAAAGAT 1050
Db 1252 AA 1311
OY 1051 GAGGACAGCAGAGAAGAGATCAAGAGATCGCAAAACCTTAAGTTCACATAGATTCT 1110
Db 1312 AA 1371
OY 1111 CTCTTACTGATCCCTCTTGAGTGTGAGTACTTCTTGAGAGAGATTAAGATATAGAC 1170
Db 1372 AA 1431
OY 1171 ATCTCCGCCAAAGTCCAGACAAAGAAATCAACGACCTTAATATATCCCATTTGGTGTG 1230
Db 1432 AA 1491
OY 1231 ACCTACCTCTGCTTATACGATATCAACAGCGCTCTCAAGAGCTCAATAGCTTCGGT 1290
Db 1492 AA 1351
OY 1291 GACTTGATTACCCCTTTCGTTTATACGAAAGAACCTCTAAGATATCTACACAGACAAT 1350
Db 1552 AA 1611
OY 1351 GAGAGAAAGAGATTATTCACGAAATCAAGAGAGATCAAAATTTGAGAGAGAAAT 1410
Db 1612 AA 1671
OY 1411 GAGAGTGACAGAAAGATTTCGAGAGACCGCAGAAAGTCTAAGCATATCACTAAGAG 1470
Db 1672 AA 1731
OY 1471 TATGAAGAGCTGCTGAACGAGATCTATGATTCGAATTCACATTAACATGACGACCTGACC 1530
Db 1732 NAAA 1791
OY 1531 AACCTGAGAAATGATGGGAAACGCTACTTACAAAGTGAGAAATGACACACAT 1590
Db 1792 AA 1851
OY 1591 AATA 1594
Db 1852 AAAA 1855

RESULT 45
BH040136/c
LOCUS
DEFINITION RPCI-24-255A23.TV RPCI-24 Mus musculus genomic clone RPCI-24-255A23
DNA sequence.
ACCESSION BH040136
VERSION BH040136.1 GI:14818806
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 691)
REFERENCE
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintet,B., Levins,M.,
Tsagaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
JOURNAL
COMMENT Other GSSs: RPCI-24-255A23.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

RESULT 47
AZ548467 908 bp DNA linear GSS 14-NOV-2000
LOCUS ENTKEK30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic DNA sequence.
ACCESSION AZ548467 GI:11172102
VERSION 1
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 908)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HML:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HML:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 17
High quality sequence stop: 828.
Location/Qualifiers
1..908
/organism="Entamoeba histolytica"
/strain="HML:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999)."

BASE COUNT 434 a 16 c 289 g 169 t
ORIGIN

Query Match 1.4%; Score 68.6; DB 17; Length 908;
Best Local Similarity 45.1%; Pred. No. 0.0001;
Matches 343; Conservative 0; Mismatches 409; Indels 9; Gaps 2;

QY 1338 CTACACAGACATGACAGAAAGATTTCACGAATCAGAGAGATCAAAATTTGA 1397
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 7 CTGGAAGAGAGAACAGAGATGATGAAGATGATGATGAAGAGATGATGAAG 66
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1398 GAAGAAGAAATTCAGAGTACAGAAAGTTTCGAAGCCGACGAAAGTCAACGA 1457
111111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 67 GAAGATGATGATGAGAGAGATGATGATGATGAAGAGATGATGAAGAGATGA 126
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1458 TATCATTAAGAGATGAAAGCTCTGAACGATCTATGATTCCAATTCACAAATTA 1517
1
DB 127 GAAGAAGATGATGAAGATGATGAAGAGATGATGAAGAGATGATGAAGATGA 186
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1518 CATGACCTGACCACTTCGAGAAATGATGGAAAAGCGT---ACTCTTACAAGTGA 1574
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 187 GATGATGATGATGAAGAGAGATGATGATGATGATGATGATGATGATGATGA 246
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1575 GAACGTGACACACCAATACCTTTCATCTTGAGATTTCAACATTAATTTGAGAA 1634
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

DB 247 GAATATGATTAAGAGATGATGATGATGAAGAGACATGATGATGAAGAGAA 306
QY 1635 GCTACACCAAGCTCTTACGATATGAGAGACTATTTCTCGGACACTTGTGGAGAA 1694
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 307 GATGATGATGAAGATGATGATGAAGAGAGATGAAGAGATGATGAAGAGATGA 366
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1695 AGAATTAAGTATTAAGAGATTCATTAAGATCGAAAACGATCGACG----- 1749
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 367 GATGATGATGAAGAGAGATGATGATGATGAAGAGATGATGAAGAGATGAAG 426
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1750 -CTTGTGAGACATTAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1808
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 427 TATGATTAAGAGATGATGATGAAGAGAGAGATGATGATGAAGAGAGATGAT 486
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1809 CGAAATTAACAGATGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 1868
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 487 GATGAAGATGATGATGAAGAGAGAGATGATGATGATGATGATGATGATGATGAT 546
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1869 GAAGGTGCTCTGATGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 1928
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 547 GATGAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 606
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1929 CGTGAAGTTAAACATTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 1988
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 607 GATGAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1989 ACCATACCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2048
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 667 GATGATGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 726
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 2049 AGTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2089
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 727 GATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 767
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

RESULT 48
BJ420171 544 bp mRNA linear EST 10-MAR-2002
LOCUS BJ420171 Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION BJ420171 Dictyostelium discoideum cDNA clone ddv38p08 5', mRNA sequence.
ACCESSION BJ420171
VERSION BJ420171.1 GI:19332411
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 544)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1..544
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="ddv38p08"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/sex="mat A"
/dev_stage="growth phase"
BASE COUNT 364 a 39 c 94 g 47 t
ORIGIN

Query Match 1.4%; Score 68.4; DB 13; Length 544;
Best Local Similarity 47.9%; Pred. No. 0.0001;
Matches 229; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

TITLE	Human gene number estimate provided by genome wide analysis using
JOURNAL	Tetraodon nigroviridis DNA sequence
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 907) Roest-Crollius H., Jalllon, O., Dasilva, C., Fizames, C., Fisher, C., Bonneu, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 907)
AUTHORS	Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (12-APR-2000) This sequence is a single read and was generated as part of a large scale clone and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .

SOURCE	BASE COUNT
1. 90/ /organism="Tetradodon nigroviridis" /db_xref="taxon:99883" /clone="225N09" /clone_lib="C" /note="Genoscope sequence ID : C0AG225CG05P1-end : PUC-Or1"	27 a 30 c 22 g 759 t 69 others

Query Match	1.4%	Score 68	DB 17	Length 907
Best Local Similarity	39.2%	Pred. NO	0.00014	
Matches 337	Conservative 41	Mismatches 482	Indels 0	Gaps 0

[illegible]

QY	1285	TGATTAACCCCTTCATTATACGAAGAACCCTTAAGAAATATCTACACAGACAAATGAGA	1354
Db	268	AAA	209
QY	1355	GAAGAACTTTATACACAAATCAAGAGAGAGATCAAAATTTGAGAGAGAAATTTGACA	1414
Db	208	AAA	149
QY	1415	GTGACAAAGAAAGTATGACGAGCCGACAGCAAAAGTCTAAAGATATCATTAAGAGTATG	1474
Db	148	AAA	89
QY	1475	AAAAGCTCTGAGACGAGATCTATGATTCCAAAATTCACAAATTAACATTCGACCTGACCAACT	1534
Db	88	AAA	29
QY	1535	TCGAGAAATGATGGGAAAA	1554
Db	28	AAAAAAAAAAAAAAAAAAAAA 9	

RESULT	51
LOCUS	Bj360881
DEFINITION	Bj360881 494 bp mRNA linear EST 07-MAR-2002
ACCESSION	Bj360881 Dictyostelium discoideum cDNA library, CF Dictyostelium
VERSION	Bj360881.1 GI:19260476

SOURCE ORGANISM	Source
Dictyostellium discoideum.	1
Dictyostellium discoideum.	2
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.	3

AUTHORS TITLE JOURNAL COMMENT	unpublished (2002)
Urushihara, H., Tanaka, Y., Kohara, Y., and Shin-I, T. Full length cDNA of Dictyostelium discoidium at the culmination stage	Contact: Tadasu Shin-i Center for Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshin@genes.nig.ac.jp.

FEATURES	source	location/Qualifiers
	1..494	/organism="Dictyostellium discoideum"
		/strain="AX4"
		/db_xref="taxon:44689"
		/clone="ddc8k12"
		/clone_lib="Dictyostellium discoideum cdna library, CF"
		/sex="mat A"
		/dev_stage="Culmination stage"
BASE COUNT	279 a	129 c 34 g 48 t 4 others
ORIGIN		

Query Match	1.48;	Score 67.2;	DB 13;	Length 494;
Best Local Similarity	50.58;	Pred. No. 0.00019;		
Matches 162; Conservative	0;	Mismatches 159;	Indels 0;	Gaps 0;

[illegible]

QY	1463	CTAAGAGTATGAAAAGCTGCTGAACGAGATCTATGATTCCTCAATTCACATTAACATCG	1522
Db	368	ACAACACAAACAGCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA	427
QY	1523	ACCTGACCACTCGAGAAA	1543
Db	428	ACAACACAAACAAACAAACAA	448
RESULT 52			
LOCUS	AU087665	500 bp	linear
DEFINITION	AU087665 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA clone XPfn5428 similar to Plasmodium falciparum surface antigen p190 gene fragment, mRNA sequence.		EST 27-JAN-2001
ACCESSION	AU087665		
VERSION	AU087665.1	GI:12389806	
KEYWORDS	EST.		
ORGANISM	Plasmodium falciparum 3D7.		
SOURCE	Plasmodium falciparum 3D7		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 500)		
TITLE	Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.		
JOURNAL	FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum		
MEDLINE	Nucleic Acids Res. 29 (1), 70-71 (2001)		
COMMENT	20574754		
FEATURES	Location/Qualifiers		
source	1..500		
	/organism="Plasmodium falciparum 3D7"		
	/strain="3D7"		
	/db_xref="taxon:36329"		
	/clone="XPfn5428"		
	/clone_lib="Sugano Malaria cDNA library"		
	/dev_stage="erythrocytic stage"		
BASE COUNT	226 a 61 c 55 g 148 t 10 others		
ORIGIN			
Query Match	1.4%;	Score 67.2;	DB 9; Length 500;
Best Local Similarity	51.5%;	Pred. No. 0.00019;	
Matches 238;	Conservative 0;	Mismatches 205;	Indels 19; Gaps 3;
QY	1065	AGAGATCAAGAGATGCGCAAAACCATTAAGTTCACATAGATTCCTCTTACTGATCC	1124
Db	39	AAAAATAAATAAAATGGCCANAACTATTAATTAATTAATGATGATTAATTAATGATGATCC	98
QY	1125	CGTTGAGCTGGAGTACTCTGAGAGAGAAAGATTAAGATATAGACATCTCCGCCAAGT	1184
Db	99	ACTGTAATTAAGAAATATTTTAAAGGAAAAAATTAATAAAGTTGATGTAACACCTAAATTC	158
QY	1185	CGA-----GACAAAGATCAACCGAACCTTAATGAATCCCATGGTGAGTGA	1235
Db	159	ACAAGATCTCTAGGAATCTGTTCAAAATACCAAAAGTCTCTATTCCTCAAAATGGTATTTGTA	218
QY	1236	CGCTCTGCTTATATACGATATCAACA-----CGCTCTCAACGAGCTCAATAGCTT	1286
Db	219	TGCTTTACACAGTACGATGATTTCAATATTCATTTAGCTGCAGATATGATTAATAATTCATA	278
QY	1287	CGGTACTGATTAACCCCTTCGATTTATAGAAAGAACCCCTCTAAGATATATCTACACAGA	1346

DB	279	TGGTATTTTAATGATTCCTCACTAAGAAAGAAAAATTAATGAAAAAATTTATTACAGATTA	338
OY	1347	CATATGAGAAAGAAAGCTTTATCAACGAATACAGAGAAGATCAAAATTTGAGAGAGAA	1406
DB	339	TAAAGAAAGAAAGAAATTTTCATTTAAATTAACATTTAAAAAATTTGATTTGAGAGAAAAAAC	398
OY	1407	AATGAGAGTGCAAGAAAGATTTACGAGACCGCAGCAAAAGTCTAAACGATPACATCA	1466
DB	399	ATTAAATCAACAAAAAGAACACATATAAAAAATTTACTTGAGATTATGAA-AGTCNAAAAA	457
OY	1467	AGAGTATGAAAAAGCTGTGAACGAGATCATGATTTCCAATTT	1508
DB	458	GGATTATGAGAAATTTACTTGAAAAATTTTATGAATGAATTT	499
RESULT 53			
LOCUS	CNS02156		
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 22AF10 of library G from Tetraodon nigroviridis, genomic survey sequence.		
ACCESSION	CNS02156	827 bp	DNA
VERSION	ALI76451		
KEYWORDS	ALI76451.1 GI:7814508		
SOURCE	GSS: genome survey sequence.		
ORGANISM	Tetraodon nigroviridis.		
REFERENCE	Tetraodon nigroviridis		
AUTHORS	Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Tetraodon.		
TITLE	1 (bases 1 to 827)		
JOURNAL	Roeest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.		
REFERENCE	2 (bases 1 to 827)		
AUTHORS	Roeest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.		
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 827)		
AUTHORS	Genoscope.		
JOURNAL	Submitted (12-APR-2000)		
COMMENT	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .		
FEATURES	location/Qualifiers		
source	1..827		
	/organism="Tetraodon nigroviridis"		
	/db_xref="taxon:99883"		
	/clone="22AF10"		
	/clone_lib="G"		
	/note="Genoscope sequence ID : C0AG224DC05DP1-end ; PUC-ori"		
BASE COUNT	368 a 203 c 106 g 121 t 29 others		
ORIGIN			
Query Match	1.3%: Score 66.4; DB 17; Length 827;		
Best Local Similarity	43.2%: Pred. No. 0.00032;		
Matches 223; Conservative 16; Mismatches 277; Indels 0; Gaps 0;			
OY	1249	AACGATATCAACACGCTCTCAACGAGCTCATATAGCTTGGTGACTTGATTAACCCCTTC	1308
DB	209	AATACACAAATTAACACACACACACAAATATCAACACATATACAGCATATACACAAAT	268
OY	1309	GATTATACGAAAGAACCTCTTATAGATATTTCTACACAGACATGACAGAAAGATTATC	1368

